9

PATENT ABSTRACTS OF JAPAN

E5874

(11)Publication number : 07-200191
(43)Date of publication of application : 04.08.1995

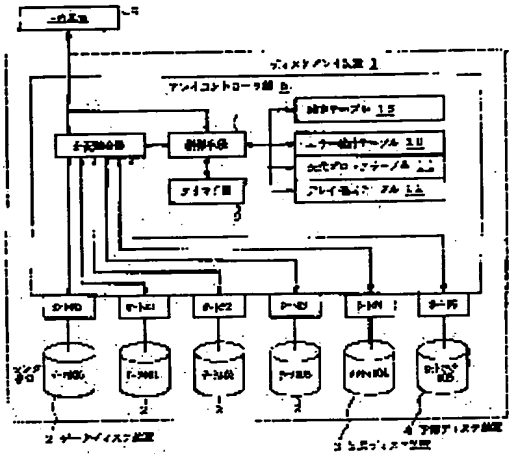
(51)Int.Cl. G06F 3/06
G06F 3/06
G06F 3/06

(21)Application number : 06-000722
(22)Date of filing : 10.01.1994

(71)Applicant : FUJITSU LTD
(72)Inventor : SATO KEIICHI

(54) DISK ARRAY DEVICE

(57)Abstract:
PURPOSE: To quickly recognize the need for data restoration to a stand-by disk device and restore data without delaying processing from a host device
CONSTITUTION: The disk array device 1 is equipped with plural disk devices 2, and at least one redundant disk device 3 and the stand-by disk device 4. A timer means 9 measures the time of data arrival from the respective data disk devices 2 and when the data arrival is delayed exceeding a specific time, a control means 7 restores delayed data from remaining disk devices. If a data block which can not be corrected is detected on the disk device 2 or 3, data are restored from the remaining disk devices which belong to the same parity group and allocated to an alternative area. Further, data are restored to the stand-by device as to a disk device which is high in the possibility of an alternative area overflow, rotation synchronism trouble, and a fault.



LEGAL STATUS

[Date of request for examination] 10.08.2000
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

JP#09

E5874

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（A）**　(11)特許出願公開番号

特開平7－200191

(43)公開日　平成７年(1995)８月４日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ０６Ｆ　3/06	５４０			
	３０５　Ｃ			
	３０６　Ｂ			

審査請求　未請求　請求項の数13　ＯＬ　（全 32 頁）

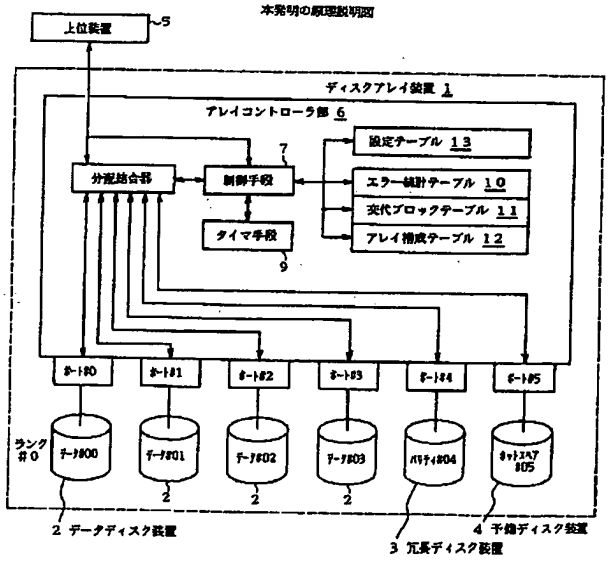
(21)出願番号　特願平6－722

(22)出願日　平成６年(1994)１月10日

(71)出願人　000005223
富士通株式会社
神奈川県川崎市中原区上小田中1015番地
(72)発明者　佐藤　恵一
山形県東根市大字東根元東根字大森5400番２　株式会社山形富士通内
(74)代理人　弁理士　竹内　進　（外１名）

(54)【発明の名称】　ディスクアレイ装置

(57)【要約】
【目的】予備ディスク装置に対するデータ復元の必要性をいち早く認識し、また上位装置からの処理を遅延することなしにデータ復元を可能とする。
【構成】ディスクアレイ装置１は、複数のデータディスク装置２、少なくとも１台の冗長ディスク装置３及び予備ディスク装置４を備える。各データディスク装置２からのデータ到着時間をタイマ手段９で計測し、所定時間を越えて遅延した場合に、制御手段７で残りのディスク装置から遅延しているデータ復元する。ディスク装置２，３のずれかで訂正不可能なデータブロックが検出された場合、同一パリティグループに属する残りのディスク装置からデータ復元して交代領域に割り付ける。更に、交代領域オーバフロー、回転同期障害、故障の危険性の高いディスク装置につき、予備ディスク装置へのデータ復元を行う。



【特許請求の範囲】

【請求項１】複数のデータディスク装置（２）、少なくとも１台の冗長ディスク装置（３）及び予備ディスク装置（４）を備えたディスクアレイ装置に於いて、

配下の各データディスク装置（２）からのデータ到着時間を監視するタイマ手段（９）と、

該タイマ手段（９）で監視しているデータ到着時間が所定時間を越えて遅延した場合に、遅延したディスク装置のデータを残りのディスク装置から復元する制御手段（７）と、を設けたことを特徴とするディスクアレイ装置。

【請求項２】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下の各データディスク装置（２）で訂正不可能なデータブロックが検出された場合に、訂正不可能となったデータを残りのディスク装置から復元し訂正不可能なデータブロックを検出したディスク装置の交代領域に割り付けることを特徴とするディスクアレイ装置。

【請求項３】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下のデータディスク装置（２）及び冗長ディスク装置（３）のいずれかで交代領域のブロックが全て使用された場合に、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、予備ディスク装置（４）に対する復元終了まで又は復元中に再度交代割付処理が要求されるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項４】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下のデータディスク装置（２）及び冗長ディスク装置（３）のいずれかで使用可能な交代領域内ブロック数が閾値まで減少した場合に、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、データ復元終了まで又は復元中に再度交代割付処理が要求されるまで又は交代ブロックが全て使用されるまで、当該ディスク装置を論理ディスク装置の構成要素として前記上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項５】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、回転同期制御を行っている配下のデータディスク装置（２）及び冗長ディスク装置（３）のいずれかの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、データ復元終了まで又は復元中にライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成用として前記上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項６】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、回転同期制御を行っている配下のデータディスク装置（２）及び冗長ディスク装置（３）のいずれかの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として処理し、更に、データ復元中に上位装置（５）よりライト命令を受けた場合にも、回転同期を維持することのできなくなったディスク装置を論理ディスク装置の構成要素として前記上位装置（５）からのライト動作を行わせることを特徴とするディスクアレイ装置。

【請求項７】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下の各ディスク装置のエラー回数とその頻度を管理するエラー統計テーブル（１０）を有し、該エラー統計テーブル（１０）を参照して故障の危険性の高いディスク装置を認識することを特徴とするディスクアレイ装置。

【請求項８】請求項７記載のディスクアレイ装置において、前記制御手段（７）は、前記エラー統計テーブル（１０）を参照して故障の危険性の高いディスク装置を認識した場合に、故障する以前に該ディスク装置のデータを先行して前記予備ディスク装置（４）に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として前記上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項９】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、データ復元終了まで又はデータ復元中に前記上位装置（５）から交代ブロック領域のデータブロックに対するライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として前記上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項１０】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、回転同期制御を行っている配下の各ディスク装置の同一トラック又は同一シリンダ位置に設けた交代領域の使用済みブロック数を管理する交代ブロックテーブル（１１）を有し、前記交代ブロックテーブル（１１）で管理される交代ブロック数が閾値を越え交代領域割付けのシーク動作に伴いデータ到達が遅延するディスク装置を認識した場合、該ディスク装置のデータを予備ディスク装置（４）に復元すると共に、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として前記上位装置（５）からの入出力要求を処理することを特徴とするディスクアレイ装置。

【請求項１１】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル（１２）を有し、前記アレイ構成テーブル（１２）を参照して配下のディスク装置のデータが予

備ディスク装置（４）に復元されていることを認識した場合に、前記予備ディスク装置（４）と故障ディスク装置との物理的位置を交換移動を促すメッセージを外部に出力して再構成させることを特徴とするディスクアレイ装置。

【請求項１２】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル（１２）を有し、前記アレイ構成テーブル（１２）を参照して配下のディスク装置のデータが予備ディスク装置（４）に復元されていることを認識した場合に、前記予備ディスク装置（４）を故障ディスク装置との論理的位置を移動させ、物理的に移動することなく再構成させることを特徴とするディスクアレイ装置。

【請求項１３】請求項１記載のディスクアレイ装置において、前記制御手段（７）は、前記上位装置（５）からの命令で、データディクス装置（２）及び冗長ディスク装置（３）のデータを前記予備ディスク装置（４）に復元させることを特徴とするディスクアレイ装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、アレイ状に配置された複数の複数のディスク装置を並列的にアクセスするディスクアレイ装置に関し、特に、ディスク装置の故障に対し予備ディスク装置を利用してデータ復元を行うディスクアレイ装置に関する。現在、電子計算機の外部記憶装置には、主に磁気ディスクが使用されている。

【０００２】近年、電子計算機の性能向上はめざましく、外部記憶装置にも高性能のものが求められている。特に、画像処理や科学技術計算の分野では、磁気ディスク装置を並列に配置し、複数台の磁気ディスクから同時にデータを読み書きしてデータ転送を高速化したディスクアレイ装置が用いられている。また冗長ディスク装置を装備し、データディスク装置に障害が起きても残りのディスク装置からデータを復元することが可能であるが、障害に対し迅速に対応可能な信頼性の高いディスクアレイ装置が求められている。

【０００３】

【従来の技術】従来、ディスクアレイ装置は小型のディスク装置を多量に使用し、システムの故障率を下げるために一般的に冗長性を持たせている。図２１は従来のディスクアレイ装置を示す。ディスクアレイ装置１はアレイコントローラ部６と複数のデータディスク装置２－０～２－３、冗長ディスク装置３およびホットスペアとして知られた予備ディスク装置４で構成される。各ディスク装置２－０～２－３，３，４はアレイコントローラ部６に設けたポート１５－０～１５－５に接続され、並列的に動作できる。アレイコントローラ部６には、プロセッサ８と分配結合器８が設けられる。

【０００４】上位装置５からのライトコマンドの発行に伴って転送されてきたブロックデータは、アレイコントローラ部６の分配結合器８で、４台のデータディスク装置２－０～２－３のデータに分配（ストライピング）され、ポート１５－０～１５－３を介して並列的に書込まれる。同時に４つの分配データからパリティデータが生成され、ポート１５－４を介して冗長ディスク装置３に書込まれる。

【０００５】ここでブロックデータはビット又はバイト単位に分配され、また冗長ディスク装置の位置を固定していることから、いわゆるＲＡＩＤ３に従った制御を行っている。データのリード時には、ライト時と逆にデータディスク装置２－０～２－３からデータを読み出してアレイコントローラ部６の分配結合器８で結合して元のブロックデータを復元し、同時に冗長ディスク装置３から読み出したパリティデータと比較して正常に結合されている場合に、復元したブロックデータを上位装置５に転送する。

【０００６】またディスクアレイ装置１では、仮に１台のデータディスク装置、例えばデータディスク装置２－３が故障した場合、残りのデータディスク装置２－０～２－２のデータと冗長ディスク装置３のパリティデータから故障したデータディスク装置２－３のデータを復元することができる。このため予備ディスク装置４が設けられ、例えばディスク装置２－３で致命的な故障が起きた場合、残りのデータディスク装置２－０～２０－２のデータと冗長ディスク装置３のパリティデータから故障したデータディスク装置２－３のデータを復元して予備ディスク装置４に書込み、予備ディスク装置４の復元データを使用可能とする。

【０００７】

【発明が解決しようとする課題】しかしながら、このような従来のディスクアレイ装置にあっては、配下のディスク装置に訂正不可能な媒体エラーや致命的なハードウェアエラーが起きた場合にのみ、故障と判断して予備ディスク装置に対するデータ復元を行っており、復元処理を行っている間、上位装置からアクセスができない不都合があった。

【０００８】また、故障したディスク装置のデータを予備ディスク装置４に復元した後、故障ディスク装置を正常なものと交換し、交換したディスク装置に予備ディスク装置４から復元データをコピーした後に正常な運用に移行させていたため、データ復元やディスク装置の交換に手間と時間がかかるという問題があった。本発明は、このような従来の問題点に鑑みてなされたもので、予備ディスク装置に対するデータ復元の必要性をいち早く認識し、また上位装置からの処理を遅延することなしにデータ復元を可能とする冗長性が高く予備ディスク装置を効率良く使用可能なディスクアレイ装置を提供することを目的とする。

【０００９】

【問題点を解決するための手段】図１は本発明の原理説明図である。まず本発明は、複数のデータディスク装置２、少なくとも１台の冗長ディスク装置３及び予備ディスク装置４を備えたディスクアレイ装置１を対象とする。

［タイマ監視によるデータ復元処理］このようなディスクアレイ装置１につき本発明は、配下の各データディスク装置２からのデータ到着時間をタイマ手段９で計測し、データ到着時間が所定時間を越えて遅延した場合に、制御手段７で遅延したディスク装置のデータを残りのディスク装置から復元する。

【００１０】尚、冗長ディスク装置３からのデータが遅延しても、データディスク装置２のデータのみからホストデータは生成できるため、冗長ディスク装置３のパリティデータを復元する必要はない。

［データ復元を伴う交代割付処理］またディスクアレイ装置１の制御手段７は、配下の各データディスク装置２および冗長ディスク装置３のいずれかで訂正不可能なデータブロックが検出された場合に、訂正不可能となったデータを同一パリティグループに属する残りのディスク装置から復元し、訂正不可能なデータブロックを検出したディスク装置の交代領域に割り付ける。

【００１１】［交代領域オーバフローに伴なう予備データ復元処理］さらにディスクアレイ装置１の制御手段７は、交代領域が全て使用済みとなった場合にも予備ディスク装置へのデータ復元を行う。即ち、配下のデータディスク装置２及び冗長ディスク装置３のいずれかで交代領域が全て使用された場合に、そのディスク装置のデータを予備ディスク装置４に復元する。

【００１２】この場合、予備ディスク装置４に対する復元終了まで、又は、データ復元中に再度交代割付処理が要求されるまでは、当該ディスク装置を論理ディスク装置の構成として上位装置５からの入出力要求を処理する。その後に全交代領域が使用済みとなったディスク装置は、故障ディスク装置として扱われ、論理ディスク装置の構成要素から除外される。

【００１３】［交代領域オーバフロー予測に基づく予備データ復元処理］交代領域の使用済みに基づく予備ディスク装置４へのデータ復元処理は、予測的に行ってもよい。即ち、データディスク装置２および冗長ディスク装置３のいずれかで使用可能な交代領域内ブロック数（空き数）が閾値まで減少した場合に、当該ディスク装置のデータを予備ディスク装置４に復元する。

【００１４】この場合、データ復元終了まで、又はデータ復元中に再度交代処理が要求されるまで又は交代ブロックが全て使用されるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置５からの入出力要求を処理する。その後は、制御手段７において故障ディスク装置として扱われる。

［同期回転異常時の予備データ復元処理］一方、ディスクアレイ装置１の制御手段７は、配下のデータディスク装置２及び冗長ディスク装置３のいずれかでの同期回転異常を検出した場合に、該ディスク装置のデータを予備ディスク４に復元する。

【００１５】この場合、データ復元終了まで又はデータ復元中にライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素として上位装置５からの入出力要求を処理する。その後は、制御手段７において故障ディスク装置として扱われる。この場合、データ復元中に上位装置５よりライト命令を受けても故障ディスク装置とせず、回転同期を維持することのできなくなったディスク装置をそのまま使用してライト動作を行わせるようにしてもよい。

【００１６】［障害発生の予測処理］ディスクアレイ装置１の制御手段７は、ディスク装置の障害発生を予測して予備ディスク装置４に対するデータ復元を行う。即ち、制御手段７は、配下の各ディスク装置のエラー回数とその頻度を管理するエラー統計テーブル１０を有し、エラー統計テーブル１０を参照して故障の危険性の高いディスク装置を認識する。

【００１７】［障害発生の予測に基づく予備データ復元処理］エラー統計テーブル１０を参照して故障の危険性の高いディスク装置を認識した場合、故障する以前に当該ディスク装置のデータを先行して予備ディスク装置４に復元する。この場合、データ復元終了まで当該ディスク装置を論理ディスク装置の構成要素として上位装置５からの入出力要求を処理する。

【００１８】データ復元が済むと、故障する危険性の高いディスク装置は、制御手段７により故障ディスク装置として扱われる。

［交代ブロックの増加に伴うディスク装置の予備データ復元処理］またディスアレイ装置の制御手段７は、回転同期制御を行っている配下の各ディスク装置の同一トラック又は同一シリンダの交代領域の使用状態（交代ブロック数）を管理する交代ブロックテーブル１１を有し、交代ブロック数が閾値に達して別のトラック又はシリンダに交代領域が割付られ、このような交代領域へのデータ割付けに伴うシーク動作でデータ到着時間が遅延する。このようにデータ到達の遅延が予測されるディスク装置のデータを予備ディスク装置４に復元する。

【００１９】この場合、データ復元終了まで、又はデータ復元中に上位装置５から交代ブロック領域のデータブロックを含むライト命令を受けるまで、当該ディスク装置を論理ディスク装置の構成要素としてとして上位装置５からの入出力要求を処理する。また、データ復元中に上位装置５から交代領域のデータブロックを含むライト命令を受けても当該ディスク装置を故障ディスク装置として除外せず、データ復元終了まで論理ディスク装置の構成要素として上位装置５からの入出力要求の処理を継続させてもよい。

【００２０】［人的介入を伴なう再構成処理］ディスクアレイ装置１の制御手段７は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル１２を有し、アレイ構成テーブル１２を参照して配下のディスク装置のデータが予備ディスク装置４に復元されていることを認識した場合に、予備ディスク装置４と故障ディスク装置との物理的位置を交換移動を促すメッセージを例えば上位装置５に送って外部に出力し、ユーザや保守要員に再構成させる。

【００２１】［人的介入を伴なわない再構成処理］ディスクアレイ装置１の制御手段７は、配下の各ディスク装置の固有情報、動作状態、物理的な設置位置を管理するアレイ構成テーブル１２を参照してデータが予備ディスク装置４に復元されていることを認識した場合に、予備ディスク装置４を故障ディスク装置の論理的位置に移動させ、物理的に移動することなく再構成させる。

【００２２】

【作用】このような本発明のディスクアレイ装置にあっては、次の作用が得られる。

［タイマ監視によるデータ復元処理］タイマ手段でデータ到達時間の監視を行い、故障を起こした１台のデータディスク装置のディスク転送が規定時間以上遅れた時、残りのデータディスク装置と冗長ディスク装置から遅延したデータディスク装置のデータを復元することによって、遅延しているデータディスク装置が各種リトライ処理を行っているような場合でも、その結果を待つことなしに上位装置に対しホストデータを転送することができ、上位装置に対する転送速度が向上する。

【００２３】［データ復元を伴う交代割付処理］訂正不可能なデータブロックと同一のパリティグループを構成する残りのディスク装置のデータから、訂正不可能なブロックのデータを復元して交代領域に割付けることで、訂正不可能ブロックを検出したディスク装置を故障ディスクとして扱わなくとも良いため、ディスク装置を有効に利用することができる。

【００２４】［交代領域オーバフローに伴なう予備データ復元処理］交代領域が使用済みとなっているディスク装置のデータを予備ディスク装置に復元し、故障ディスク装置と見做して交換可能とする。この場合、データ復元終了まで、又は、このディスク装置においてデータ復元中に再度交代割付処理が必要となるまで継続して使用し続けることによって、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、論理ディスク装置の冗長性を確保でき、ユーザデータを保証することができる。

【００２５】［交代領域オーバフロー予測に基づく予備データ復元処理］一方、交代領域の使用済みに伴う予備ディスク装置のデータ復元中に再度交代割付処理が必要となった場合、ディスク装置の冗長性が失われてしまうのに対し、交代領域の使用可能な残りブロック数が閾値以下に減少した時に予備ディスク装置へのデータ復元を開始することで、テータ復元中にも残り数分の交代割付処理かできるので、より高い信頼性でユーザデータを保証することができる。

【００２６】また、データ復元中に起こり得る交代割付処理以上に残りブロック数の閾値を上げておくことで、データ復元後まで論理ディスク装置の冗長性を保つことができる。

［同期回転異常時の予備データ復元処理］回転同期が維持できなくなったディスク装置のデータを予備ディスク装置に復元し、故障ディスク装置として交換可能とする。

【００２７】この場合、同期回転が維持できなくなったディスク装置を、データ復元終了まで、又は上位装置からライト命令を受け取るまで継続して使用し続けることによって、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、ディスク装置の冗長性を確保でき、ユーザデータを保証することができる。また、データ復元中の上位装置５からのライト命令を高速に処理できる。

【００２８】一方、データ復元中にライト命令に対し、同期回転が維持きなくなっているディスク装置にライト動作を行わせてもよい。これにより回転同期を維持している時よりライト動作に時間はかかるが、回転同期を維持することが不可能となったディスク装置を、データ復元終了まで継続して使用し続けて冗長性が確保され、より高い信頼性でユーザデータを保証することができる。

【００２９】［障害発生の予測処理］エラー統計テーブルの内容を設定されている閾値と比較することによって、故障する危険性の高いディスク装置を事前に把握できる。

［障害発生の予測に基づく予備データ復元処理］故障する危険性の高いディスク装置が致命的な故障を起こす以前に予備ディスク装置にデータを復元することができる。また、データ復元中に他のデータディスク装置または冗長ディスク装置が故障した場合にも、ディスクの冗長性を確保できるためユーザデータを保証することができる。

【００３０】［交代ブロック数に伴うディスク装置の予備データ復元処理］同一のトラック又は同一シリンダの交代領域の交代ブロック数が増加すると他のトラック又はシリンダに交代領域が拡張され、このような交代ブロックをリードするとシーク動作に時間がかかり、上位装置へのデータ転送速度が遅延する。このようなディスク装置のデータを予備ディスク装置に復元し、交代領域を使用せずにリード可能な状態にすることで、データ転送の遅延を抑えることができる。

【００３１】［人的介入を伴なう再構成処理］何らかの原因により配下のディスク装置のデータが予備ディスクに復元されている場合に、予備ディスク装置のデータを

交換したディスク装置にコピーすること無しに、ユーザや保守要員が物理的な位置を移動するだけで論理ディスク装置を再構成して使用することができる。
［人的介入を伴なわない再構成処理］予備ディスク装置の物理的な位置を移動せずに、論理ディスク装置を再構成して使用でき、ディスク装置の交換に伴う人為的ミスを防止できる。
【００３２】
【実施例】
＜目　次＞
１．ハードウェア構成と機能
２．タイマ監視によるデータ復元
３．データ復元を伴う交代割付処理
４．交代領域オーバフローに伴う予備へのデータ復元
５．交代領域オーバフロー予測に基づく予備へのデータ復元
６．同期回転異常時の予備へのデータ復元
７．障害発生の予測と予備へのデータ復元
８．同一トラック内の交代ブロック数の増加に伴う予備へのデータ復元
９．人的介入を伴う再構成
１０．人的介入を伴わない再構成
１．ハードウェア構成と機能
図２は本発明のディスクアレイ装置のハードウェア構成と機能を示した実施例構成図である。
【００３３】図２において、本発明のディスクアレイ装置１は例えば４台のデータディスク装置２－０～２－３、１台の冗長ディスク装置３、および１台の予備ディスク装置４を備え、各ディスク装置をアレイコントロール部６に設けられたポート１５－０～１５－５のそれぞれに接続している。ディスクアレイ装置１における各ディスク装置は、その論理的位置および物理的位置がポート番号＃０～＃５とランク番号＃０～＃ｎで特定される。この実施例にあっては、１ランク構成を例にとっているが、必要に応じて複数ランク設けることができる。
【００３４】この実施例にあっては、ポート番号＃０～＃３のポート１５－０～１５－３にデータディスク装置２－０～２－３を接続し、またポート番号＃４のポート１５－４にパリティデータを格納する冗長ディスク装置３を接続している。データディスク装置２－０～２－３および冗長ディスク装置３のアドレス（デバイスＩＤ）はポート番号とランク番号で特定され、＃００，＃０１，＃０２，＃０３，＃０４となる。
【００３５】４台のデータディスク装置２－０～２－３と１台の冗長ディスク装置３により、１つのパリティグループを構成している。例えば、上位装置５からのデータブロックを書き込む場合には、転送されたデータブロックをアレイコントローラ部６で４つのデータディスク装置２－０～２－３に分配するストライピング処理を行い、同時に各ストライピングデータからパリティデータを生成し、各ストライピングデータおよびパリティデータを並列的にポート１５－０～１５－４からデータディスク装置２－１～２－３および冗長ディスク装置３に供給して、並列的に書き込む。
【００３６】一方、上位装置５からのリード要求に対しては、アレイコントローラ部６でリードブロックのアドレスからデータディスク装置２－０～２－３の各アドレスを生成し、更に冗長ディスク装置３についてもアドレスを生成し、ストライピングデータおよびパリティデータを系列的にリードする。そしてアレイコントローラ部６において、リードした各ストライピングデータからリードブロックを生成し、生成したリードブロックからパリティを生成して冗長ディスク３からリードしたパリティデータとの整合をチェックし、正常であれば上位装置５にリードブロックを転送することになる。
【００３７】このようなディスクアレイ装置１におけるライト動作およびリード動作のため、アレイコントローラ部６には制御手段としてのプロセッサ７および分配結合器８が設けられている。なお、この実施例にあっては、冗長ディスク装置３を固定的に決めていることから、データブロックをビット単位あるいはバイト単位に分配して系列的に読み書きする、いわゆるＲＡＩＤ３として知られたディスクアレイの制御形式を例にとっている。
【００３８】更に本発明のアレイコントローラ部６にあっては、タイマ９，エラー統計テーブル１０，交代ブロックテーブル１１，アレイ構成テーブル１２および閾値設定テーブル１３を新たに設けている。プロセッサ７はタイマ９，エラー統計テーブル１０，交代ブロックテーブルテーブル１１，アレイ構成テーブル１２および閾値設定テーブル１３を使用し、配下のデータディスク装置２－０～２－３および冗長ディスク装置３のいずれかで障害が発生したことを認識すると、障害発生ディスク装置を除く他のディスク装置からのデータに基づく復元処理、あるいは障害と判定されたディスク装置のデータを予備ディスク装置４に復元する復元処理を実行する。
【００３９】即ち、本発明によるプロセッサ７の制御処理は、次に列挙する内容となる。
①タイマ９の監視によるリードブロックデータの復元
②データ復元を伴う交代割付処理
③交代領域がオーバフローになった場合の予備ディスク装置４へのデータ復元
④交代領域のオーバフローを予測して、予備ディスク装置４へのデータ復元
⑤回転同期制御を行っている各ディスク装置のいずれかにおける同期回転異常の際の予備ディスク装置４へのデータ復元
⑥各ディスク装置の障害の発生を予測して、予備ディスク装置４へのデータ復元
⑦交代ブロック数が増加したディスク装置の予備ディス

ク装置へのデータ復元
⑧予備ディスク装置４へのデータ復元が済んだ後の人的介入による再構成
⑨予備ディスク装置への復元が済んだ後の、人的介入を伴わない再構成
以下、前記①～⑨のそれぞれについて詳細に説明する。
２．タイマ監視によるデータ復元
図３は上位装置５からのリード命令に対し、アレイコントローラ部６のプロセッサ７が適当なタイミングでタイマ９を初期化して起動し、パリティグループを構成するデータディスク装置２－０～２－３のいずれか１台からのデータ転送が閾値設定テーブル１３に設定されているタイムアウト時間以上遅れた場合、既に得られている残りのデータディスク装置からのデータに基づき、遅延したデータディスク装置のデータを復元し、リードブロックを生成して上位装置５に転送するようにしたことを特徴とする。
【００４０】図３のフローチャートについて詳細に説明すると次のようになる。まずステップＳ１で、ホストコンピュータ５よりプロセッサ７がリード要求を受領すると、ステップＳ２で、ホストコンピュータにおける論理ディスクを構成する配下のデータディスク装置２－０～２－３および冗長ディスク装置３に対しリードコマンドを発行する。
【００４１】このリードコマンドの発行に基づきステップＳ３で、配下のデータディスク装置２－０～２－３および冗長ディスク装置３のリード処理が行われ、リードデータが転送されてくる。このときプロセッサ７はどのディスク装置が分配結合器８に対しデータ転送を完了したか否かを監視しており、パリティデータとデータ復元に必要なリードデータが分配結合器８に転送されるか否か、ステップＳ４でチェックしている。
【００４２】即ち、この実施例にあっては、４台のデータディスク装置２－０～２－３を配下にもつことから、パリティデータと４台のディスク装置のうちの３台のリードデータが得られると、残り１つのリードデータを復元することができるため、４台のデータディスク装置２－０～２－３のうちの３台のデータ転送と冗長ディスク装置３からのパリティデータの転送が済んだ時点で、ステップＳ４のデータ復元に必要なデータ受領と判定する。
【００４３】続いてステップＳ５のデータ遅延のフラグセットをチェックするが、初期状態でフラグはリセットされていることからステップＳ６に進み、プロセッサ７はタイマ９の初期化および起動を行う。ステップＳ６におけるタイマ９の起動後、プロセッサ７は閾値設定テーブル１３より、予め定められたタイムアウト時間（閾値時間）を読み出す。
【００４４】ステップＳ８で、設定されたタイムアウト時間内に遅延しているディスク装置からのデータを受領した場合には、分配結合器８で全てのデータディスク装置２－０～２－３からのリードデータからホストコンピュータ５に対するリードブロックデータを生成し、冗長ディスク装置３からのパリティデータとの整合性を確認した上で、ホストコンピュータ５に対しリードデータから生成されたホストブロックデータを転送する。
【００４５】一方、ステップＳ８の設定タイムアウト時間内に残りのデータディスク装置からのリードデータを受け取ることができなかった場合には、ステップＳ９に進んで、データ遅延が発生したことを示すフラグをセットし、既に分配結合器８に転送されている３台のデータディスク装置からのリードデータと冗長ディスク装置３からのパリティデータを使用して、ステップＳ１０でホストデータブロックを生成してホストコンピュータ５に転送する。
【００４６】続いてプロセッサ７は、データ転送が遅延しているデータディスク装置をその後も監視し、遅延の原因と結果を示すステータス情報の受領を待つ。このステップＳ１１におけるステータス情報の受領待ちの際に、ステップＳ１２で再度リード要求があるか否かチェックしている。もしステータス情報を受領する前に再度リード要求があると、ステップＳ２に戻り、ホストコンピュータからのリード要求に基づきステップＳ２～Ｓ３で配下のデータディスク装置へのリードコマンドの発行に伴うリード処理およびデータ転送を行わせ、この場合にも前回と同様、同じデータディスク装置からのデータ遅延を起こすことになる。
【００４７】この場合、前回の処理で既にデータ遅延フラグがセットされていることから、ステップＳ５から直ちにステップＳ１０に進み、ステップＳ６～Ｓ９のタイマ起動に基づく監視処理は行わず、ステップＳ４で分配結合器８に得られた遅延ディスク装置を除くデータディスク装置からのリードデータと冗長ディスク装置３からのパリティデータによってデータを復元してホストブロックデータを生成して、ホストコンピュータ５に直ちに転送する。
【００４８】ステップＳ１１で、遅延しているディスク装置の原因と結果を示すステータス情報が受領されると、ステップＳ１３に進み、データ遅延を示すフラグをクリアし、ステップＳ１４で、データ遅延の原因が各種のリトライ処理にあったか否かチェックする。遅延理由が各種のリトライ処理にあった場合には、続いてステップＳ１５で、データ遅延の原因が訂正不可能なリードデータのリトライ処理であったか否かチェックする。
【００４９】もし訂正不可能なリードデータのリトライ処理であった場合には、図４のステップＳ１６以降の処理に進む。ここで、訂正不可能なリードデータのリトライ処理は主にディスク媒体の媒体欠陥を原因とするものであり、交代領域に対する割付処理でリカバーすることができる。一方、ステップＳ１５でデータ遅延の原因が

訂正不可能なリードデータのリトライ処理でなかった場合、即ちデータディスク装置のエラーであった場合には、図５のステップＳ２１に示すエラー統計テーブル１０の更新処理に移行する。

３．データ復元を伴う交代割付処理

図４のフローチャートは、図３のステップＳ１５に示したデータ遅延を起こしたデータディスク装置について訂正不可能ブロックが検出された場合の交代割付処理を示している。

【００５０】図４において、ステップＳ１～Ｓ１５の処理は図３と同じであり、ステップＳ２～Ｓ１４については省略して示している。ステップＳ１５でデータ遅延の原因が訂正不可能なデータブロックの検出にあったことが判別された場合、ステップＳ１６に進み、訂正不可能なデータブロックのアドレスから他の正常なデータディスク装置での同一パリティグループを構成するアドレスを算出し、次のステップＳ１７で、正常なデータディスク装置２および冗長ディスク装置３から訂正不可能なデータブロックが検出されたディスク装置のデータ復元に必要なデータリードのためのリードコマンドを発行する。

【００５１】続いてステップＳ１８で、正常なデータディスク装置２および冗長ディスク装置３から得られたリードデータおよびパリティデータに基づき、訂正不可能なブロックから検出されたディスク装置のリードデータを復元し、ステップＳ１９で、データ遅延を起こしたデータディスク装置の交代領域に復元したデータを書き込む割付けを行う。

【００５２】このようなステップＳ１６～Ｓ１９における訂正不可能なデータブロックを他のディスク装置からのリードデータで復元して交代領域に割り付ける処理は、ディスクアレイ装置１がホストコンピュータ５からの命令を何も実行していない空き時間に行うことが望ましい。またステップＳ１９で、交代領域に対するデータ割付けが済んだ場合には、図５のステップＳ２１およびＳ２２に示すエラー統計テーブル１０の更新および交代ブロックテーブル１１の更新を行うようになる。

４．交代領域オーバフローに伴う予備へのデータ復元

図５のフローチャートは、配下のディスク装置において交代割付用の交代領域が全て使い尽くされたときに、交代領域を全て使い尽くしたディスク装置のデータを予備ディスク装置４に復元するようにしたことを特徴とする。

【００５３】具体的に図５について説明すると次のようになる。なお図５のフローチャートにあっては、データディスク装置２－３（機番アドレス＃０３）で交代領域が使い尽くされてオーバフローとなった場合の処理を示している。尚、図のフローチャート中でデータディスク装置は、データ＃０３と省略して示している。まずステップＳ２０で、配下のディスク装置のうちのデータディスク装置２－３が訂正不可能なブロックの検出に基づき、交代領域への割付処理を実施したとする。続いてステップＳ２１で、交代処理の原因となったエラーのエラー統計テーブル１０に対する更新を行い、またステップＳ２２で、交代ブロックテーブル１１に対する更新を行う。

【００５４】次にステップＳ２３でデータディスク装置２－３に関する交代ブロックテーブル１１の内容を参照し、交代領域の残りブロック数が０となって使い尽されていた場合には、ステップＳ２４に進み、プロセッサ７は直ちに交代領域が使い尽くされたデータディスク装置２－３のデータを予備ディスク装置４に復元するための処理を開始する。

【００５５】このデータ復元処理は、交代領域が使い尽くされたデータディスク装置２－３を除く他のデータディスク装置２－０～２－２および冗長ディスク装置３のリード処理を行って、分配結合器８でデータディスク装置２－３のデータを復元し、これを予備ディスク装置４に書き込む処理を行う。ステップＳ２４で予備ディスク装置４へのデータ復元を開始すると、データ復元中におけるホストコンピュータ５からのリード要求に対してはステップＳ２５～Ｓ３０に示す処理が行われ、一方、ライト要求についてはステップＳ３２～Ｓ３８（Ｓ４１，Ｓ４２を含む）の処理が行われる。

【００５６】まずホストコンピュータ５からデータ復元中にリード要求を受けると、このリード要求はステップＳ２５で判別され、ステップＳ２６に進む。ステップＳ２６にあっては、データディスク装置２－３が故障ディスク装置に設定されているか否かチェックする。最初、データディスク装置２－３は故障ディスク装置には設定されていないことから、ステップＳ２７に進み、データディスク装置２－３における交代割付処理が必要なリード要求か否かチェックする。

【００５７】ステップＳ２７でデータディスク装置２－３の交代割付処理を必要としないリード要求であった場合には、ステップＳ２８に進み、全てのデータディスク装置２－０～２－３および冗長ディスク装置３に対しリードコマンドを発行してリード動作を行わせ、分配結合器８で各データデータディスク装置２－０～２－３のリードデータからホストブロックデータを生成し、パリティデータとの整合をチェックした後、ホストコンピュータ５に転送する。

【００５８】一方、ステップＳ２７でデータディスク装置２－３に関し、交代割付処理を必要とするリード要求であった場合には、データディスク装置２－３について交代処理を行ってリードデータを転送すると処理時間がかかることから、ステップＳ２９でデータディスク装置２－３を故障ディスク装置に設定し、ステップＳ３０でディスク装置２－３を除くデータディスク装置２－０～２－２および冗長ディスク装置３に対しリードコマンド

を発行して、得られたデータからホストブロックデータを復元してホストコンピュータ５に転送する。

【００５９】次に図６のステップＳ３１で予備ディスク装置４へのデータ復元中にホストコンピュータ５からのライト要求が判別されると、ステップＳ３２に進み、データディスク装置２－３は故障ディスク装置に設定されているか否かチェックする。故障ディスク装置に設定されていなければステップＳ３３に進み、データディスク装置２－３の交代割付処理を伴うライト要求か否かチェックする。交代割付処理を伴うライト要求でなければ、ステップＳ３４で全てのデータディスク装置２－０～２－３および冗長ディスク装置３に対しライト命令を実行する。

【００６０】一方、ステップＳ３３でデータディスク装置２－３に関し交代割付処理を必要とするライト要求であった場合には、データディスク装置２－３の交代領域は使い尽くされて使用不可となっていることから、ステップＳ３５に進み、データディスク装置２－３を故障ディスク装置に設定する。

【００６１】続いてステップＳ３６で、故障ディスク装置に設定されたデータディスク装置２－３を除くデータディスク装置２－０～２－２および冗長ディスク装置３に対しライト命令を実行する。この場合、故障ディスク装置に設定されたデータディスク装置２－３分の分配データ（ストライピングデータ）が欠落するが、他のディスク装置のライトデータおよびパリティデータから復元可能であることから問題はない。

【００６２】なお、予備ディスク装置４に対する復元済みのデータについてのライト要求であった場合には、ライト要求の実行完了後に再度、予備ディスク装置４への復元データに割り付ける必要がある。ステップＳ３１のライト要求に伴う処理が済むと、ステップＳ３７で、予備ディスク装置４への復元を行っているデータディスク装置２－３以外のデータディスク装置２－０～２－２に障害発生か否かをチェックした後、ステップＳ３８で予備ディスク装置４へのデータ復元終了を監視し、予備ディスク装置４へのデータ復元終了まで以上の処理を繰り返す。

【００６３】ステップＳ３８で予備ディスク装置へのデータ復元終了が判別されると、ステップＳ３９で、データ復元中にデータディスク装置２－３は故障ディスク装置に設定されたか否かチェックし、もし設定されていなければステップＳ４０でデータディスク装置２－３を故障ディスク装置とする。一方、データ復元中のステップＳ３７で、復元対象となっているデータディスク装置２－３以外のデータディスク装置２－０～２－２または冗長ディスク装置３に障害が発生した場合には、ステップＳ４１に進んで、データディスク装置２－３が故障ディスク装置に設定されているか否かチェックする。

【００６４】もし故障ディスク装置に設定されていた場合にはパリティグループに属する２台のディスク装置に故障が起きていることから、この場合には冗長性が失われ、ユーザデータの崩壊につながり、異常終了となる。ステップＳ４０でデータディスク装置２－３がまだ故障ディスク装置として設定されていない場合には、ステップＳ４２に進み、データディスク装置２－３について行っていた予備ディスク装置４へのデータ復元を中止し、新たに障害を起こしたディスク装置のデータを予備ディスク装置４に対し行うようにする。即ち、交代領域が使い尽くされた場合の予備ディスク装置４へのデータ復元に対し、致命的な障害となるような故障ディスク装置についての予備ディスク装置４へのデータ復元処理を優先させる。

【００６５】図５および図６に示した交代領域オーバフローに伴う予備ディスク装置へのデータ復元処理にあっては、交代領域が使い尽くされたデータ復元の対象となっているデータディスク装置２－３の故障ディスク装置への設定は、もしリード要求またはライト要求により交代割付処理を必要としなければ、予備ディスク装置４へのデータ復元終了まで交代領域が使い尽くされたデータディスク装置２－３を正常なディスク装置としてリード動作またはライト動作することができ、冗長性および並列アクセス性能が確保できる。

【００６６】一方、データ復元中に割付交代処理を必要とするリード要求またはライト要求があると、交代領域が使い尽くされたデータディスク装置２－３は故障ディスク装置に設定され、リード動作およびライト動作の対象から除外される。しかしリード動作については、他のディスク装置のリードデータから故障ディスク装置に設定されたデータディスク装置２－３のリードデータを復元できることから、リード性能の低下は起きない。

【００６７】またライト動作についても、故障ディスク装置として設定されたデータディスク装置２－３に対するライト動作が行われないだけであることから、ライト性能も低下することはない。

５．交代領域オーバフロー予測に基づく予備へのデータ復元

図７および図８のフローチャートは、ディスク装置の交代領域の残り数が予め定めた閾値に減少したときに予備ディスク装置へのデータ復元を行うようにしたことを特徴とする。

【００６８】図７において、ステップ２０で例えばデータディスク装置２－３が交代割付処理を実施したとすると、ステップＳ２１でエラー統計テーブル１０の更新を行い、またステップＳ２２で交代ブロックテーブル１１の更新を行う。次にステップＳ２３で、閾値設定テーブル１３を参照して、予め定めたデータ復元処理を開始するための交代領域の残りブロック数の閾値を求め、ステップＳ２４で、交代ブロックテーブル１１から求めたデータディスク装置２－３の現在の交代領域残りブロック

数と比較する。

【００６９】この交代領域残りブロック数が閾値と等しければ、ステップＳ２５に進み、データディスク装置２－３のデータを予備ディスク装置４に復元する処理を開始する。予備ディスク装置４へのデータ復元中にホストコンピュータ５からリード要求があると、ステップＳ２６～Ｓ３２の処理が行われ、一方、ライト要求があると、図８のステップＳ３３～Ｓ４０の処理が行われる。

【００７０】図７のデータ復元中のリード要求については、ステップＳ２６でリード要求が判別されてステップＳ２７に進み、データディスク装置２－３が故障ディスク装置に設定されていなければ、ステップＳ２８で交代割付処理を必要とするか否かチェックし、必要としなければステップＳ２９で、パリティグループに含まれる全ディスク装置を対象にリード命令を実行する。

【００７１】ステップＳ２８で、交代領域残りブロック数が閾値に減少したデータディスク装置２－３に関し交代割付処理が必要となった場合には、ステップＳ３０で、交代領域残りブロック数が０に達したか否かチェックし、達していなければ、ステップＳ３１の故障ディスク装置への設定を行わず、ステップＳ３２で、データディスク装置２－３以外のパリティグループに属するディスク装置でリード命令を実行する。

【００７２】即ち、この場合の予備ディスク装置へのデータ復元の開始は閾値分の空きブロックが残っている状態で開始していることから、交代領域残りブロック数が０になるまではデータディスク装置２－３を故障ディスク装置とはせず、冗長性を確保する。データ復元中のライト要求は図８のステップＳ３３で判別され、ステップＳ３４に進み、交代領域残りブロック数が閾値に減少したデータディスク装置２－３は故障ディスク装置に設定されたか否かチェックし、設定されていなければ、ステップＳ３６で交代割付処理を必要とするライト要求か否かチェックし、必要としなければステップＳ３７で、パリティグループを構成する全ディスク装置に対しライト命令を実行する。

【００７３】ステップＳ３６でデータディスク装置２－３に関し交代割付処理を必要とするライト要求であった場合には、ステップＳ３８で交代領域残りブロック数が０か否かチェックし、０に達するまではステップＳ３９でデータディスク装置２－３を故障ディスク装置に設定せず、ステップＳ４０でデータディスク装置２－３以外のディスク装置でライト命令を実行する。

【００７４】このようなライト要求に対するデータ復元中の処理が済むと、ステップＳ４１でデータディスク装置２－３以外のディスク装置の障害発生の有無をチェックした後、ステップＳ４０で予備ディスク装置４へのデータ復元終了の有無をチェックし、これを繰り返す。ステップＳ４２で予備ディスク装置４へのデータ復元終了が判別されると、ステップＳ４３で、データ復元中にデータディスク装置２－３は故障ディスク装置に設定されたか否かチェックし、設定されていなければステップＳ４４で故障ディスク装置に設定して、一連の処理を終了する。

【００７５】一方、ステップＳ４１でデータ復元中にデータディスク装置２－３以外のディスク装置で障害が発生した場合には、ステップＳ４５に進み、このときデータディスク装置２－３が故障ディスク装置に設定されていなければ、ステップＳ４６で予備ディスク装置４へのデータ復元を中断し、新たに故障した他のディスク装置のデータの予備ディスク装置４への復元を開始する。

【００７６】またステップＳ４５で、データ復元中に既にデータディスク装置２－３が故障ディスク装置に設定されていた場合には、同一パリティグループに含まれる２台のディスク装置に故障が起きたことから冗長性が失われ、ユーザデータの崩壊によって異常終了となる。

６．同期回転異常時の予備へのデータ復元

図２に示した本発明のディスクアレイ装置１にあっては、アレイコントローラ部６の配下に設けたデータディスク装置２－０～２－３，冗長ディスク装置３および予備ディスク装置４は、スピンドルモータによるディスク媒体の回転制御につき回転同期をとる制御を行う場合がある。このように各ディスク装置のスピンドルモータの回転同期制御が行われている場合には、特定のディスク装置に回転同期の異常が起きると、異常を起こしたディスク装置のアクセス性能の低下に伴って全体的な性能低下を引き起こすことになる。

【００７７】そこで本発明にあっては、図９のフローチャートに示すように回転同期の維持ができなくなったことを検出して、そのディスク装置のデータを予備のディスク装置に復元する処理を行うようにしたことを特徴とする。図９において、パリティグループを構成しているディスク装置の中の特定のディスク装置、例えばデータディスク装置２－３で回転同期が維持できなくなったとき、この回転同期の異常をプロセッサ７が受領し、ステップＳ２で、回転同期が維持できなくなったデータディスク装置２－３のデータを予備ディスク装置４に復元するデータ復元処理を開始する。

【００７８】即ち、残りの回転同期が正常なデータディスク装置２－０～２－２および冗長ディスク装置３にリード命令を発行して、得られたリードデータを分配結合器８で生成して予備ディスク装置４に書き込む復元処理を開始する。予備ディスク装置４へのデータ復元中に、ステップＳ３でホストコンピュータ５からのリード要求が判別されると、ステップＳ４に進み、データディスク装置２－３以外の同一パリティグループに属するディスク装置のリード命令の実行でホストブロックデータをリードデータから生成して、ホストコンピュータ５に転送する。

【００７９】またデータ復元中にステップＳ５でホスト

コンピュータ５からのライト要求を判別すると、ステップＳ６で回転同期が維持できなくなっているデータディスク装置２－３を故障ディスク装置に設定し、データディスク装置２－３を除く同一グループに属するディスク装置でライト命令を実行する。この場合、同期回転が維持できないデータディスク装置２－３はライト命令の実行対象から除外されているため、ライト命令は同期回転を維持しているディスク装置で行われ、ライト処理を高速に行うことができる。

【００８０】リード要求またはライト要求に伴う処理が済むと、ステップＳ８で同期回転が維持できないデータディスク装置２－３以外のディスク装置に障害発生があったか否かチェックした後、ステップＳ９で予備ディスク装置４へのデータ復元終了の有無をチェックし、データ復元終了まで以上の処理を繰り返す。ステップＳ９で予備ディスク装置４へのデータ復元終了が判別されると、ステップＳ１０で、データ復元中に同期回転が維持できなくなっているデータディスク装置２－３が障害ディスク装置に設定されたか否かチェックし、もし設定されていなければ、ステップＳ１１で故障ディスク装置に設定する。

【００８１】一方、データ復元中にステップＳ８で同期回転が維持できなくなったデータディスク装置２－３以外のディスク装置に障害が発生した場合には、ステップＳ１２に進み、障害発生以前にライト命令を実行したか否かチェックし、もし障害発生以前にライト命令を実行していると、同期回転が維持できなくなったディスク装置２－３を含めて同一パリティグループに属する２台のディスク装置が故障したことから冗長性が失われ、ユーザデータの崩壊として異常終了に至る。

【００８２】一方、障害発生以前にライト命令が実行されていなければステップＳ１３に進み、現在行っている同期回転が維持できないデータディスク装置２－３に関する予備ディスク装置４へのデータ復元を中止し、新たに障害発生となった他のディスク装置のデータを予備ディスク装置４に復元する。この場合の予備ディスク装置へのデータ復元は、同期回転が維持できなくなっているデータディスク装置２－３を正常なディスク装置として扱うことから、データ復元処理に多少時間がかかることになる。更に、新たに故障したデータディスク装置の予備ディスク装置へのデータ復元が終了したならば、データ復元ができた予備ディスク装置を後から故障した障害ディスク装置と物理的に入れ替え、予備ディスク装置の位置に別の新たなディスク装置をセットして再度、同期回転が維持できていないデータディスク装置２－３のデータの復元処理を行う。

【００８３】図１０は同期回転を維持できなくなったディスク装置に対するデータ復元処理の第２実施例を示したフローチャートである。この第２実施例にあっては、データ復元中にホストコンピュータからのライト要求を受けても同期回転を維持できなくなったディスク装置を故障ディスク装置に設定せず、同期回転を維持できなくなっているディスク装置を正常なディスク装置としてライト動作を実行するようにしたことを特徴とする。

【００８４】即ち、図１０のステップＳ１～Ｓ４は図９のフローチャートと同じであるが、データ復元中のライト要求をステップＳ５で判別すると、ステップＳ６で、同期回転を維持できなくなっているデータディスク装置２－３を故障ディスク装置に設定せず、全ディスク装置でライト命令を実行する。このため、ライト要求に対するライト命令の実行時間は同期回転を維持できなくなっているデータディスク装置２－３の性能低下に依存するため、ライト処理が低速で時間がかかるようになる。しかしながら、ディスクアレイ装置１としての冗長性は維持できる。

【００８５】更に、データ復元中にステップＳ７で同期回転が維持できなくなっているデータディスク装置２－３以外のディスク装置に障害が発生したことを判別すると、ステップＳ１０に進み、現在の予備ディスク装置４へのデータ復元を中止し、新たに障害発生となったディスク装置のデータの予備ディスク装置への復元を先行して行う。

【００８６】これに対し、図９に示した実施例では、データ復元中にライト要求があった場合には、同期回転が維持できなくなったデータディスク装置２－３を故障ディスク装置に設定してしまっているため、他のディスク装置の障害発生に対し同一パリティグループに属する２台のデータディスク装置が故障となって冗長性が失われ、ユーザデータの崩壊につながっている。

【００８７】しかし、図１０の第２実施例では同期回転が維持できなくなってもデータディスク装置２－３を故障ディスク装置としていないことから、他のディスク装置の障害に対し冗長性を失うことなく予備ディスク装置へのデータ復元ができ、その後に同期回転が維持できなくなっているデータディスク装置２－３の予備ディスク装置４へのデータ復元が可能となる。

【００８８】更に、ステップＳ８で予備ディスク装置へのデータ復元の終了が判別されると、ステップＳ９で、最終的に同期回転が維持できなくなっているデータディスク装置２－３を故障ディスク装置と設定する。

７．障害発生の予測と予備へのデータ復元図２に示した本発明のディスクアレイ装置１にあっては、次のようなエラーを
検出したときにプロセッサ７にエラー報告を行う。

【００８９】①ディスク装置内部のリトライ処理によって回復されたリードエラー
②ディスク装置内部のリトライ処理によって回復されたポジショニング系のエラー
③ディスク装置内部のリトライ処理によって回復されたパリティエラー

④交代割付処理で回復されたライトエラー
⑤回復不可能なリードエラー
プロセッサ７はこのようなエラー検出に基づく報告を受領すると、エラー統計テーブル１０を作成または内容更新すると共に、閾値設定テーブル１３に予め設定されている閾値とエラー回数を比較し、エラー回数が閾値を越えているディスク装置を故障する危険の高いディスク装置として扱う障害予測処理を行う。

【００９０】具体的には、図１１のフローチャートのステップＳ１～Ｓ６により障害発生の予測処理が行われる。まずステップＳ１で、配下のディスク装置の各種のエラーを検出して報告を受けると、ステップＳ２で、エラー統計テーブル１０の更新の必要性の有無をチェックし、必要があれば、ステップＳ３でエラー統計テーブル１０の作成または更新を行う。

【００９１】続いてステップＳ４で、閾値設定テーブル１３を参照して予め定めた閾値を求め、ステップＳ５でエラー回数と閾値を比較する。エラー回数が閾値以上であればステップＳ６に進み、そのディスク装置を故障の危険性が高いディスク装置と認識する。ステップＳ６以降にあっては、ディスク装置２－３が故障の危険性の高いディスク装置と認識された場合を例にとって具体的に説明する。

【００９２】ステップＳ６で、データディスク装置２－３が故障の危険性が高いディスク装置と認定されると、ステップＳ７に進んで、プロセッサ７は直ちに認定されたデータディスク装置２－３のデータを予備ディスク装置４に復元するデータ復元処理を開始する。データ復元中にホストコンピュータ５からのリード要求をステップＳ８で判別すると、ステップＳ９～Ｓ１４に示す処理が行われる。即ちステップＳ９で、故障の危険性が高いと認識されたデータディスク装置２－３は故障ディスク装置に設定されているか否かチェックし、設定されていなければ、ステップＳ１０でデータディスク装置２－３に各種のエラー発生があったか否かチェックし、エラー発生がなければステップＳ１１で、全ディスク装置に対しリード命令を実行させて、ホストブロックデータを転送する。

【００９３】ステップＳ１０でデータディスク装置２－３に各種のエラーが発生した場合には、ステップＳ１２で各種リカバリ処理の成功の有無をチェックし、リカバリ処理に成功すれば、ステップＳ１３での故障ディスク装置への設定を行わず、ステップＳ１４で、故障の危険性が高いと認識されているデータディスク装置２－３を除く他のディスク装置でリード命令を実行して、復元したホストブロックデータを上位装置５に転送する。ステップＳ１２で各種リカバリ処理に失敗した場合には、ステップＳ１３で、故障の危険性の高いデータディスク装置２－３は故障ディスク装置に設定され、処理対象から除外される。

【００９４】データ復元中にホストコンピュータ５からのライト要求が図１２のステップＳ１５で判別されると、ステップＳ１６～Ｓ２１の処理が行われる。即ち、ステップＳ１６で、故障の危険性が高いデータディスク装置２－３は故障ディスク装置に設定されているか否かチェックされ、設定されていなければステップＳ１７で、各種エラーが発生したか否かチェックし、発生していなければステップＳ１８で、全ディスク装置に対しライト命令を実行させる。

【００９５】ステップＳ１７で、故障の危険性が高いと認識されているデータディスク装置２－３で各種エラーが発生した場合には、ステップＳ１９で、各種リカバリ処理の成功をチェックし、成功すれば、ステップＳ２０における故障ディスク装置への設定を行わず、ステップＳ２１で、データディスク装置２－３を除く他のディスク装置でライト命令を実行する。リカバリ処理に失敗すれば、ステップＳ２０でデータディスク装置２－３を故障ディスク装置に設定する。

【００９６】データ復元中にステップＳ２２で、故障の危険性の高いデータディスク装置２－３以外のディスク装置に障害が発生すると、ステップＳ２６で、データディスク装置２－３が故障ディスク装置に設定されていなければ、ステップＳ２７で新たに障害発生となった障害ディスク装置の予備ディスク装置に対するデータ復元を先行し、その後に故障の危険性の高いデータディスク装置２－３の予備ディスク装置へのデータ復元を行う。

【００９７】ステップＳ２６で、データ復元中にデータディスク装置２－３が既に故障ディスク装置に設定されていた場合には、同一パリティグループに属する２つのディスク装置に障害が発生して冗長性が失われていることから、ユーザデータの崩壊として異常終了する。ステップＳ２３で、予備ディスク装置４に対するデータ復元の終了が判別されると、ステップＳ２４に進み、データディスク装置２－３が故障ディスク装置に設定されていなければ、ステップＳ２５で故障ディスク装置への設定を行った後、一連の処理を終了する。

８．同一トラック内の交代ブロック数の増加に伴う予備へのデータ復元
図２に示した本発明のディスクアレイ装置１にあっては、ホストコンピュータ５からのリード要求またはライト要求に対し各ディスク装置のデータブロックが媒体欠陥などにより訂正不可能なデータブロックとなった場合には、交代領域への割付処理が行われる。この割付処理の対象となる交代領域は同一トラック上または同一シリンダ上に存在するが、同一トラック上または同一シリンダ上の交代領域ブロック数が、交代処理が進んで残り数が少なくなると、別のトラックまたは別のシリンダ位置を交代領域として新たに確保するため、その後の交代領域に対するリード動作またはライト動作の際のシーク時間が長くなる。

【００９８】そこで図１３のフローチャートに示すように、同一トラックまたは同一シリンダに含まれる交代ブロック数が予め定めた閾値を越えたディスク装置については、シーク時間の増加に伴う性能低下の原因になることから、予備ディスク装置へのデータ復元を行う。これを図１３のフローチャートについて説明すると次のようになる。

【００９９】図１３において、ステップＳ１で配下のディスク装置例えばデータディスク装置２－３で交代ブロックの割付処理が行われ、これをプロセッサ７が受領したとする。プロセッサ７は交代処理の報告を受けて、ステップＳ２で交代ブロックテーブル１１の作成または更新を行い、続いてステップＳ３で、閾値設定テーブル１３を参照して同一トラックまたは同一シリンダに割付け可能な交代ブロック数の閾値を読み出す。

【０１００】次にステップＳ４で、現在の同一トラックまたと同一シリンダの交代ブロック数と閾値設定テーブル１３から読み出した閾値とを比較し、閾値を越えている場合には、このデータディスク装置２－３について、ステップＳ５において予備ディスク装置４へのデータの復元を開始する。データ復元中にホストコンピュータ５からのリード要求がステップＳ６で判別すると、ステップＳ７～Ｓ１０の処理を行う。即ちステップＳ７で、同一トラックまたは同一シリンダの交代ブロック数が閾値を越えたデータディスク装置２－３が既に故障ディスク装置に設定されているか否かチェックし、設定されていなければステップＳ８で、リード要求ブロックは交代ブロックを含んでいるか否かチェックし、含んでいなければステップＳ９で、全ディスク装置でリード命令を実行する。

【０１０１】ステップＳ８でリード要求ブロックに交代ブロックを含んでいた場合には交代ブロックのシーク動作に時間がかかることから、ステップＳ１０に進み、データディスク装置２－３以外のディスク装置でリード命令を実行し、このリード命令で得られたリードデータおよびパリティデータからホストブロックデータを復元してホストコンピュータ５にデータ転送する。

【０１０２】予備ディスク装置４に対するデータ復元中に、図１４のステップＳ１１でホストコンピュータ５からのライト要求が判別されると、ステップＳ１２～Ｓ１６の処理が行われる。即ち、ステップＳ１２で同一トラックまたは同一シリンダの交代ブロック数が閾値に達したデータディスク装置２－３は故障ディスク装置に設定されているか否かチェックし、設定されていなければステップＳ３で、ライト要求ブロックは交代ブロックを含んでいるか否かチェックする。

【０１０３】交代ブロックを含んでいなければステップＳ１４で、全ディスク装置でライト命令を実行する。一方、ステップＳ１３でライト要求ブロックにデータディスク装置２－３の交代ブロックを含んでいる場合、ステップＳ１４で同一トラックまたは同一シリンダの交代ブロック数が閾値に達したデータディスク装置２－３を故障ディスク装置に設定する。そしてステップＳ１６で、故障ディスク装置に設定したデータディスク装置２－３を除く他のディスク装置でライト命令を実行する。

【０１０４】更に、データ復元中にステップＳ１７でデータディスク装置２－３以外のディスク装置に障害が発生すると、ステップＳ２１で、現在予備ディスク装置４へのデータ復元の対象となっているデータディスク装置２－３が故障ディスク装置に設定されていないことを条件に、ステップＳ２２で、新たに障害発生となった障害ディスク装置の予備ディスク装置へのデータ復元を行う。

【０１０５】そして新たな障害ディスク装置へのデータ復元が終了すると、一旦中断したデータディスク装置２－３に関する予備ディスク装置へのデータ復元を行う。一方、データ復元中に受けたライト要求ブロックにデータディスク装置２－３の交代ブロックが含まれていた場合には、ステップＳ１５で故障ディスク装置に設定されていることから、この場合には同一パリティグループに含まれる２台のディスク装置で故障が起きたこととなり、冗長性が失われ、ユーザデータの崩壊として異常終了となる。

【０１０６】ステップＳ１８にあっては、予備ディスク装置４へのデータ復元の終了をチェックしており、データ復元を終了するとステップＳ１９で、データディスク装置２－３の故障ディスク装置への設定が済んでいることをチェックすると、ステップＳ２０で故障ディスク装置に設定した後、一連の処理を終了する。図１５は同一トラックまたは同一シリンダの交代ブロック数が閾値に達したときの予備ディスク装置へのデータ復元処理の第２実施例を示したフローチャートである。この図１５のフローチャートに示す第２実施例にあっては、予備ディスク装置へのデータ復元中にホストコンピュータから交代ブロックを含むライト要求を受けても、データ復元対象となっているデータディスク装置２－３を故障ディスク装置に設定せずに、論理ディスク装置を構成する有効なディスク装置として冗長性を確保するようにしたことを特徴とする。

【０１０７】即ち、図１５のステップＳ１～Ｓ５の処理は図１３と同じであるが、データ復元中にステップＳ６でホストコンピュータ５からのリード要求を判別した場合には、ステップＳ７で、全リード命令を実行する。またデータ復元中にステップＳ８でホストコンピュータ５からのライト要求を判別した場合には、ステップＳ９で、全ディスク装置でライト命令を実行する。

【０１０８】このため、予備ディスク装置へのデータ復元の対象となったデータディスク装置２－３は、復元終了までホストコンピュータ５の論理ディスク装置を構成する有効なディスク装置として扱われることで冗長性が

確保される。更にデータ復元中において、図１６のステップＳ１１で、データディスク装置２－３以外のディスク装置で障害が発生した場合には、ステップＳ１５で、新たに障害発生となった障害ディスク装置のデータの予備ディスク装置４への復元を行った後、一旦中断したデータディスク装置２－３に関する予備ディスク装置４へのデータ復元を行う。

【０１０９】データ復元中にステップＳ１２で予備ディスク装置４へのデータ復元の終了が判別されると、ステップＳ１３に進む。ここで初めてデータディスク装置２－３は故障ディスク装置に設定され、論理ディスク装置を構成するディスク装置の中から除外される。

９．人的介入を伴う再構成

図２に示した本発明のディスクアレイ装置１にあっては、何らかに原因により特定のディスク装置で予備ディスク装置へのデータ復元が必要となり、予備ディスク装置４へのデータ復元が終了すると、データ復元の済んだ予備のディスク装置をホストコンピュータの論理ディスク装置を構成するディスク装置に入れ替える再構成が必要となる。

【０１１０】即ち、ホストコンピュータ５から見た論理ディスク装置を構成するディスク装置の中の故障ディスク装置を取り外して、データを復元した予備ディスク装置４に置き替える処理が必要となる。図１７のフローチャートにあっては、予備ディスク装置４に対するデータ復元の終了後の故障ディスク装置との入れ替えを、ユーザや保守要員などの人的介入により行うようにしたことを特徴とする。

【０１１１】図１７のフローチャートにおいて、まずステップＳ１で、何らかの原因によりデータディスク装置２－３の予備ディスク装置４に対するデータの復元が必要なことを受領すると、ステップＳ２に進み、プロセッサ７はアレイ構成テーブル１２を参照し、その種別情報から予備ディスク装置４を選択して、ステップＳ３で予備ディスク装置４へのデータ復元を開始する。

【０１１２】ステップＳ２のデータ復元開始前のアレイ構成テーブル１２の内容は、例えば図１８（Ｂ）に示すようになる。図１８（Ｂ）において、アレイ構成テーブル１２はディスクアレイ装置１における論理ディスクを構成する複数のディスク装置をランク番号，ポート番号，動作状態を示す種別情報で示している。即ち、アレイ構成テーブル１２内において各ディスク装置は、図１８（Ａ）に示すように、「ａｂｙ」で表わされる。この内、先頭のａはランク番号、次のｂはポート番号、最後のｙは装置の動作状態即ち役割を示す種別情報である。この種別情報はデータディスク装置はＤ、冗長ディスク装置はＰ、予備ディスク装置はＨ、故障ディスク内はＦ、データ復元中ディスク装置はＲが使用される。

【０１１３】図１８（Ｂ）のデータ復元前のアレイ構成テーブル１２にあっては、ポート番号＃０～＃３のディスク装置の論理基板は００Ｄ～０３Ｄであり、データディスク装置２であることが判る。またポート番号＃４のディスク装置は論理基板０４Ｐとなって、冗長ディスク装置３となっていることが判る。更にポート番号＃５のディスク装置は論理基板０５Ｈから、予備ディスク装置となっていることが判る。

【０１１４】このような図１８（Ｂ）に示すデータ復元前のアレイ構成テーブル１２の参照により、ステップＳ３で予備ディスク装置４を認識して予備ディスク装置４へのデータ復元を開始する。予備ディスク装置４へのデータ復元中にあっては、ステップＳ４でアレイ構成テーブル１２の内容を更新する。即ち、図１８（Ｃ）に示すように、データ復元を必要とするポート番号＃３のデータディスク装置２－３の内容を「０３Ｆ」と故障ディスク装置に変更し、データ復元を行っているポート番号＃５の予備ディスク装置４の内容を「０５Ｒ」として、データ復元中のディスク装置であることを示すように更新する。

【０１１５】ステップＳ５で予備ディスク装置５へのデータ復元が終了すると、ステップＳ６で、データ復元終了に伴うアレイ構成テーブル１２の更新を行う。即ち、図１８（Ｄ）に示すように、故障ディスク装置となったポート番号＃３のディスク装置のデータの存在を消去し、ポート番号＃５の予備ディスク装置４については復元したデータ＃０３の存在を登録すると共に、「０５Ｄ」に更新してデータディスク装置となったことを示す。

【０１１６】続いてステップＳ７で、予備ディスク装置４へのデータ復元終了をホストコンピュータに通知し、ホストコンピュータ５側のディスプレイ装置などを使用して、ユーザまたは保守要員に対し予備ディスク装置を故障ディスク装置の位置に物理的に差し替える再構成の作業を促すメッセージ出力などを行う。このホストコンピュータ５におけるメッセージ出力を受けてユーザあるいは保守要員は、ステップＳ８で、故障ディスク装置となっているデータディスク装置２－３を取り外し、データ復元の済んだ予備ディスク装置４を故障ディスク装置の位置に差し替える位置の移動を行う。

【０１１７】予備ディスク装置４の故障ディスク装置の位置への移動が済むと、ステップＳ９で、プロセッサ７は移動後のアレイ構成テーブル１２の更新を図１８（Ｅ）に示すように行う。勿論、ディスクアレイ装置１から取り外された故障ディスク装置は点検修理が行われることになる。また、空きとなった予備ディスク装置４の位置には別の正常なディスク装置あるいは処理が済んだ故障ディスク装置が実装されることになる。

１０．人的介入を伴わない再構成

図１９は何らかに原因によりデータ復元が必要と判断されて、予備ディスク装置へのデータ復元が済んだ後の人的な介入によるディスク装置の移動を必要としない再構

成処理を示している。

【０１１８】図１９にあっては、ステップＳ１で、何らかの原因により例えばデータディスク装置２－３のデータ復元が必要となった場合の処理を示している。この場合、ステップＳ２でまず復元前のアレイ構成テーブル１２をプロセッサ７が参照する。この場合のアレイ構成テーブル１２は、例えば図２０（Ｂ）に示す内容を有する。図２０（Ｂ）において、アレイ構成テーブル１２上で各ディスク装置は「ａｂｙ，ａ´ｂ´」で表わされる。先頭の「ａ，ｂ」は物理的な位置を示すランク番号ａとポート番号ｂである。次のｙはデータディスクＤ，冗長ディスクＰ，予備ディスクＨ，故障ディスクＦ，データ復元中ディスクＲとなる動作状態を示す種別情報である。最後の「ａ´，ｂ´」は論理的な位置を示すランク番号ａ´とポート番号ｂ´である。

【０１１９】このことから、図２０（Ｂ）のデータ復元前のアレイ構成テーブル１２にあっては、ランク番号＃０に属する６台のディスク装置について、ポート番号＃０～＃３のディスク装置については「００Ｄ００」～「０３Ｄ０３」が登録され、またポート番号＃４のディスク装置については「０４Ｐ０４」が登録され、更にポート番号＃５のディスク装置については「０５Ｈ」が登録されており、前半の物理的な位置と後半の論理的な位置の値は共に等しく、物理的な位置と論理的な位置が１対１に対応している。

【０１２０】このようなアレイ構成テーブル１２をもつ各ディスク装置につき、ステップＳ３で予備ディスク装置４に対するデータ復元が必要となったデータディスク装置２－３のデータの復元処理を開始する。データ復元中にあっては、ステップＳ４でアレイ構成テーブルの更新を行う。この場合のアレイ構成テーブル１２は、図２０（Ｃ）に示すように、データ復元が必要となったポート番号＃３のディスク装置２－３の登録内容を「０３Ｆ０３」として故障ディスク装置に設定し、またポート番号＃５のディスク装置を「０５Ｒ０５」としてデータ復元中のディスク装置であることを示す。

【０１２１】ステップＳ５で予備ディスク装置４へのデータ復元が終了すると、ステップＳ６で予備ディスク装置４の論理的な位置の移動を行い、ステップＳ７で、この論理的な位置移動に伴うアレイ構成テーブル１２の更新を行う。即ち、図２０（Ｄ）に示すように、ポート番号＃５のデータ復元が済んだ予備ディスク装置４について「０５Ｄ０２」に更新し、ポート番号＃３の故障ディスク装置のもっていた論理的な位置を示す番号「０３」に変更する。

【０１２２】このため、それ以降のホストコンピュータ５からのリード要求およびライト要求については、ポート番号＃５に接続されているデータ復元の済んだ予備ディスク装置４が論理的な位置「０２」をもつデータディスク装置として扱われる。また故障ディスク装置となったポート番号＃３のデータディスク装置２－３は論理的な位置から除外され、故障ディスク装置を取り外し、新たに正常なディスク装置を実装した際、アレイ構成テーブル１２のポート番号＃３の位置に予備ディスク装置としての「０３Ｈ０５」の登録が行われ、異なった物理的な位置で予備ディスク装置４として機能するようになる。

【０１２３】勿論、図１９によってはステップＳ７でデータ復元終了後のアレイ構成テーブルの更新が済むと、ステップＳ８でデータ復元終了をホストコンピュータ５に通知し、通常のホストコンピュータ５からのコマンド要求に基づく入出力処理に戻る。尚、上記の各実施例にあっては、ディスクアレイ装置１のプロセッサ７において配下のディスク装置でデータ復元が必要なことを認識した場合の予備ディスク装置に対するデータ復元の起動を行っているが、ディスクアレイ装置１からホストコンピュータ５に対しデータ復元が必要なディスク装置が認識されたことを通知し、ホストコンピュータ５からのコマンドにより予備ディスク装置４に対するデータ復元を開始するようにしてもよい。

【０１２４】この処理は例えば図５のフローチャートのステップＳ２３からＳ２４の処理に移行する部分に示すように上位装置から行われる。この点は、図７、図９、図１０、図１１、図１３、図１５、図１７及び図１９のフローチャートについても同様である。このように予備ディスク装置４に対するデータの復元をホストコンピュータ５側で管理することで、ディスクアレイ装置１側の処理負担を軽減すると同時に、ディスクアレイ装置１に対する入出力要求の空き時間を効率的に利用した予備ディスク装置４へのデータ復元処理を可能とする。

【０１２５】また本発明は１ランクに６台のディスク装置を設けた場合を例にとっているが、ランク数および１ランク当りのディスク装置の台数は必要に応じて適宜に定めることができる。また１ランクに設ける予備ディスク装置を１台としているが、複数台設けてもよい。また１ランクに１台、予備ディスク装置を設けず、複数ランク当り１台の予備ディスク装置を設けてもよい。

【０１２６】更にまた、上記の実施例はホストブロックデータをビットまたはバイト単位に分配結合して、ポートに並列接続された複数のディスク装置を並列動作し、またパリティデータを格納するディスク装置を特定ポートに固定したＲＡＩＤ３のディスクアレイ制御形態を例にとっているが、セクタ単位にデータのリード，ライトを行い，セクタごとにパリティデータを格納するディスク位置が変化するＲＡＩＤ５のディスクアレイ制御形態についても、例えばパリティグループを構成する全ディスク装置に対しセクタデータを並列的に読み書きするような場合については、そのまま適用することができる。

【０１２７】

【発明の効果】以上説明してきたように本発明によれ

ば、次の効果が得られる。まずタイマによりリード動作実行時のデータ到達時間を監視し、パリティグループの中の１台のディスク装置からのデータ転送が遅れた場合には、既に得られているデータおよびパリティデータからデータを復元して上位装置に返送することで、上位装置に対する転送速度を向上できる。

【０１２８】また訂正不可能なデータブロックについて、同一パリティグループに属する他のディスク装置のデータから訂正不可能なデータを復元して交代領域に割り付けるため、訂正ブロックを検出したディスク装置を故障ディスク装置として扱わなくてもよく、ディスク装置の信頼性を向上することができる。更に、交代領域が全て使い尽くされたり残り数が少なくなったディスク装置のデータを予備ディスク装置に復元して故障ディスク装置として扱うことで、交代領域のアクセスに伴う応答時間の遅れを防止し、ディスクアレイ装置の処理性能を維持することができる。

【０１２９】更にまた、同期回転を維持できなくなったディスク装置についても、予備ディスク装置にデータを復元して故障ディスク装置として扱うことで、同期回転が維持できなくなった装置が論理ディスク装置を構成するディスク装置の中に存在することで処理性能が低下してしまうことを防止できる。

【０１３０】更にまた、同一トラックまたは同一シリンダの交代領域ブロック数が所定値を越えた場合に、予備ディスク装置にデータを復元して故障ディスク装置として扱うことで、交代領域が複数トラックや複数シリンダ位置に及ぶことによるアクセス性能の低下を防止できる。更にまた、ディスク装置のエラー回数が閾値に達したときに予備ディスク装置へのデータ復元を行って故障ディスク装置として扱うことで、故障の危険性の高いディスク装置を論理ディスク装置を構成するディスク装置の中から除外し、信頼性を向上できる。

【０１３１】一方、何らかの原因によりデータ復元が必要となって、予備ディスク装置に対するデータ復元が終了した後の予備ディスク装置の故障ディスク装置との置き替えについて、置き替えを促すメッセージ出力を自動的に行うことで、物理的な移動を伴う人的な再構成を確実にできる。更に、予備ディスク装置の論理的な位置をアレイ構成テーブル上で書き替えるだけで、物理的な故障ディスク装置と予備ディスク装置の移動を必要とすることなく簡単にディスクアレイの再構成が実現できる。

【０１３２】

【図面の簡単な説明】

【図１】本発明の原理説明図

【図２】本発明のハードウェア構成と機能を示した実施例構成図

【図３】タイマ監視によるデータ復元処理を示したフローチャート

【図４】データ復元を伴う交代割付処理を示したフローチャト

【図５】交代領域オーバフローに伴う予備へのデータ復元処理を示したフローチャート

【図６】図５の続きを示したフローチャート

【図７】交代領域オーバフローの予測に基づくデータ復元処理を示したフローチャート

【図８】図８の続きを示したフローチャート

【図９】同期回転異常時の予備へのデータ復元処理を示したフローチャート

【図１０】同期回転異常時の予備へのデータ復元処理の他の実施例を示したフローチャート

【図１１】障害発生の予測と予備へのデータ復元処理を示したフローチャート

【図１２】図１１の続きを示したフローチャート

【図１３】同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理を示したフローチャート

【図１４】図１３の続きを示したフローチャート

【図１５】同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理の他の実施例を示したフローチャート

【図１６】図１５の続きを示したフローチャート

【図１７】人的介入により物理的なディスク装置の移動を伴う再構成処理を示したフローチャート

【図１８】図１７の処理で更新されるアレイ構成テーブルの内容を示した説明図

【図１９】人的介入により物理的なディスク装置の移動を必要としない再構成処理を示したフローチャート

【図２０】図１９の処理で更新されるアレイ構成テーブルの内容を示した説明図

【図２１】従来装置の説明図

【０１３３】

【符号の説明】

１：ディスクアレイ装置

２，２－０～２－３：データディスク装置

３：冗長ディスク装置

４：予備ディスク装置

５：上位装置（ホストコンピュータ）

６：アレイコントローラ部

７：プロセッサ（制御手段）

８：分配結合器

９：タイマ（タイマ手段）

１０：エラー統計テーブル

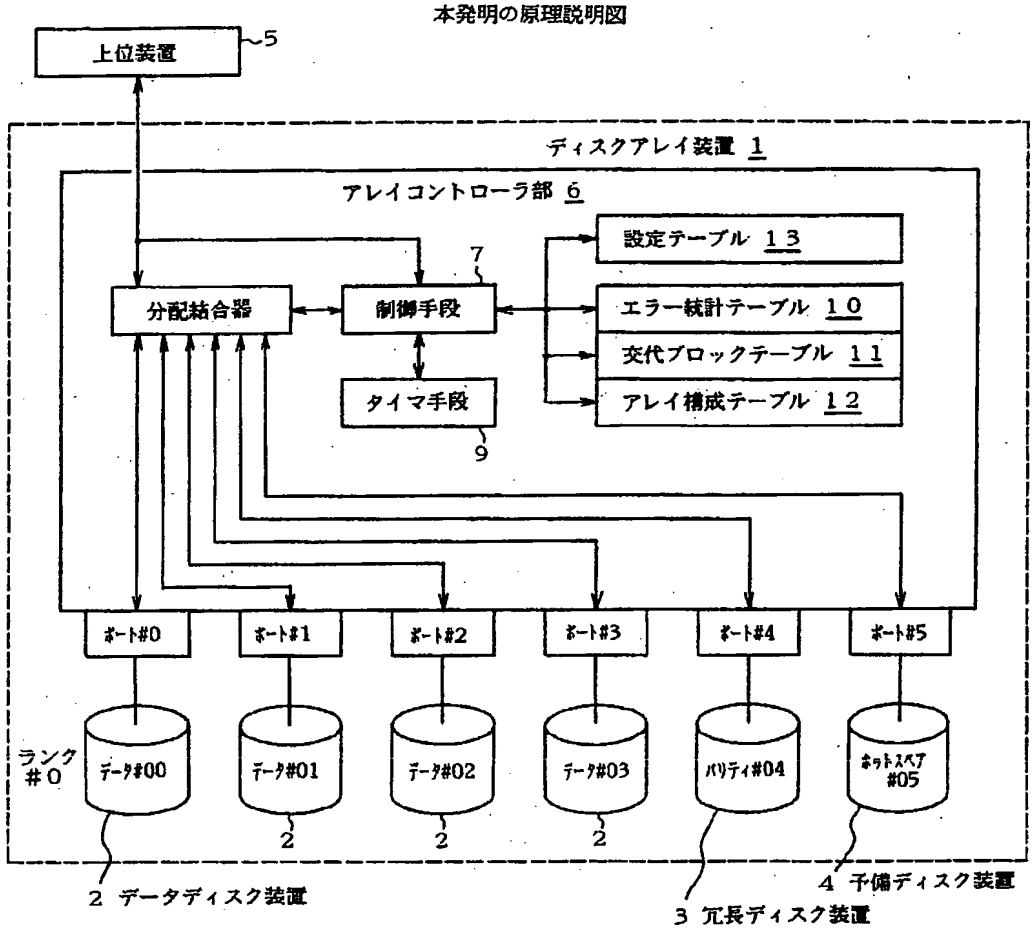
１１：交代ブロックテーブル

１２：アレイ構成テーブル

１３：閾値設定テーブル

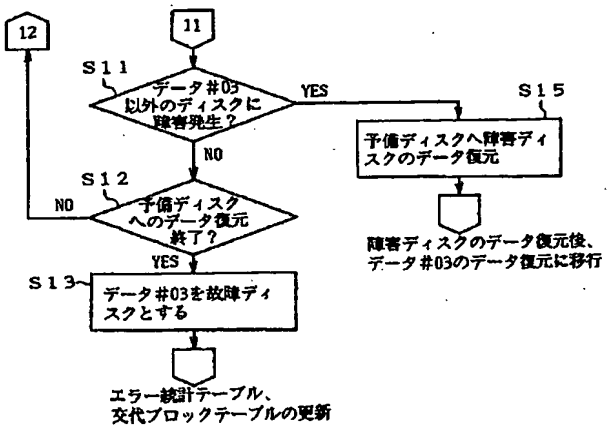
【図１】

本発明の原理説明図



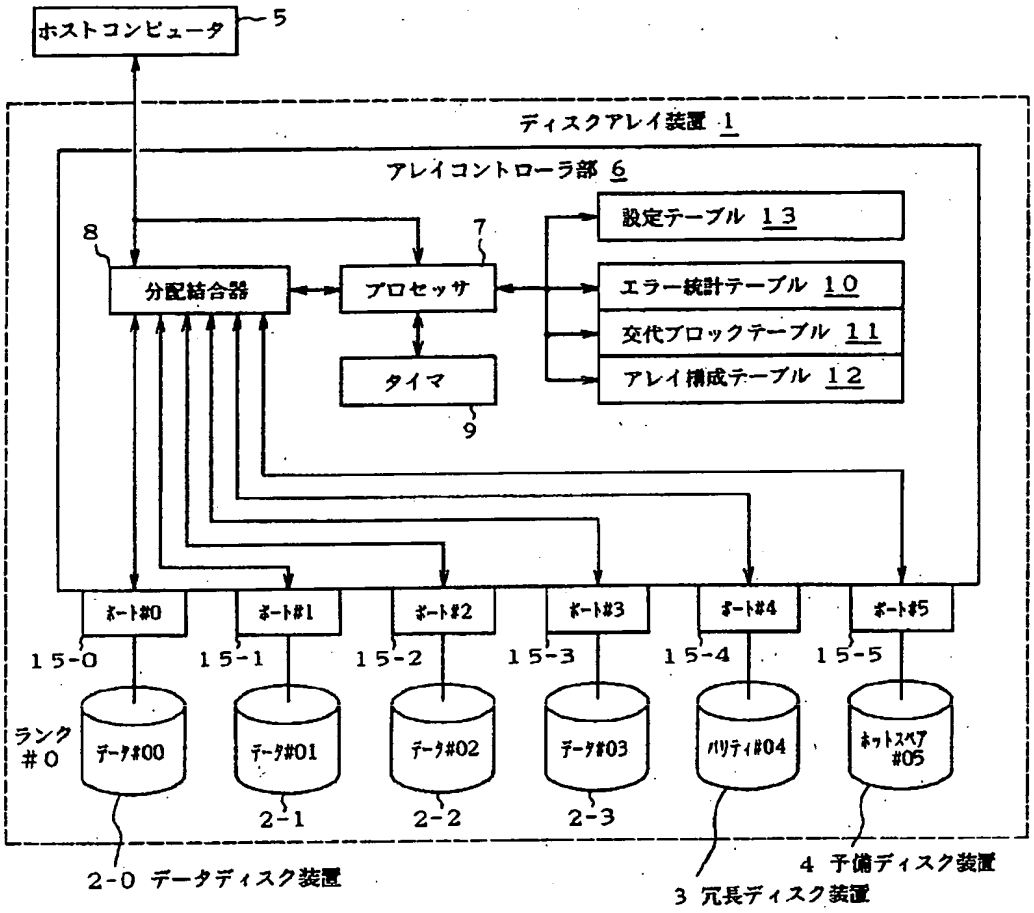
【図１６】

図１５の続きを示したフローチャート



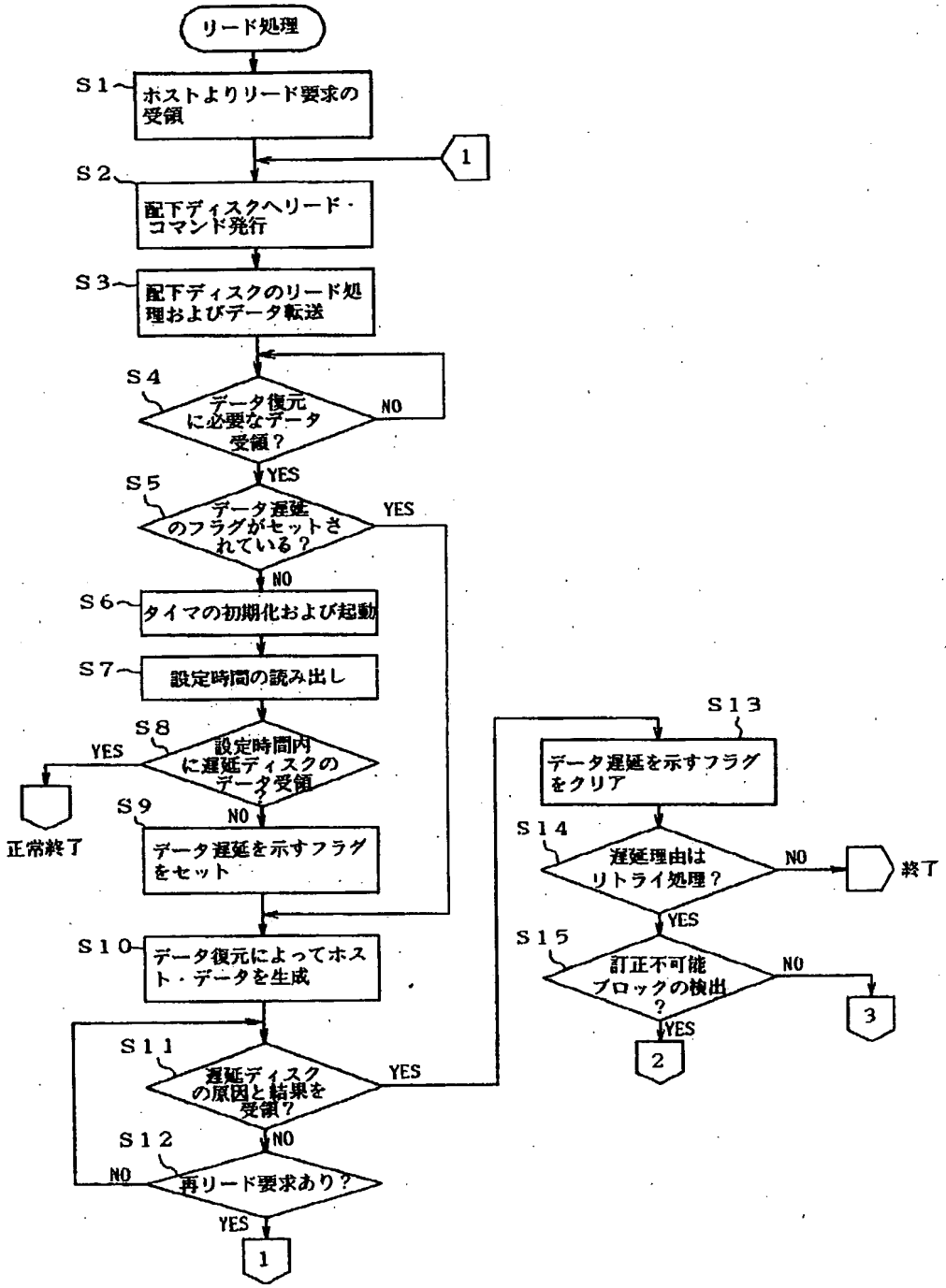
【図２】

本発明のハードウェア構成と機能を示した実施例構成図



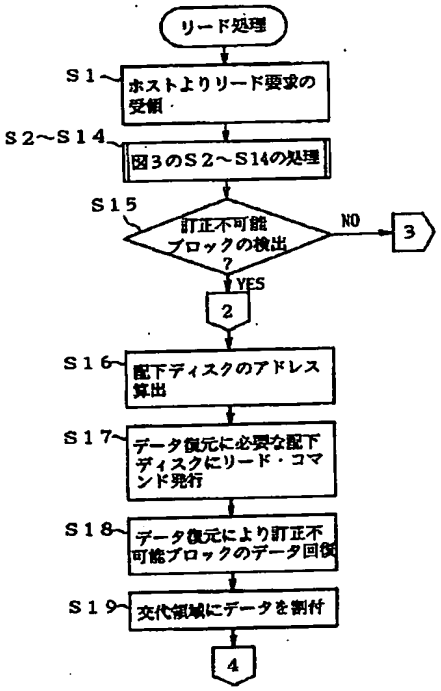
【図３】

タイマ監視によるデータ復元処理を示したフローチャート



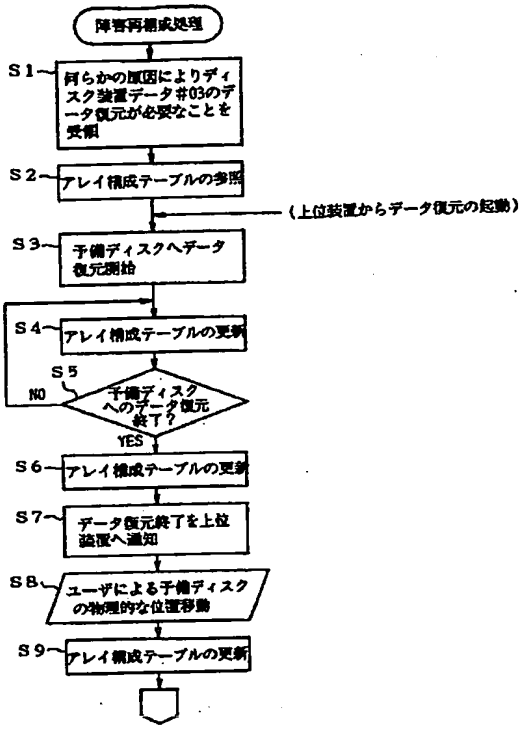
【図４】

データ復元を伴う交代割付処理を示したフローチャート



【図１７】

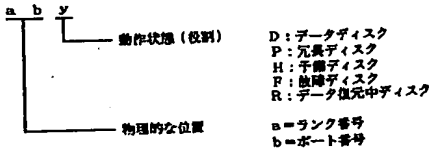
人的介入により物理的なディスク装置の移動を伴う再構成処理を示したフローチャート



【図１８】

図17の処理で更新されるアレイ構成テーブルの内容を示した説明図

(A)



(B)（データ復元前）

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	ホットスペア#0
	00D	01D	02D	03D	04P	05H

(C)（データ復元中）

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	ホットスペア#0
	00D	01D	02D	03P	04P	05R

(D)（データ復元後）

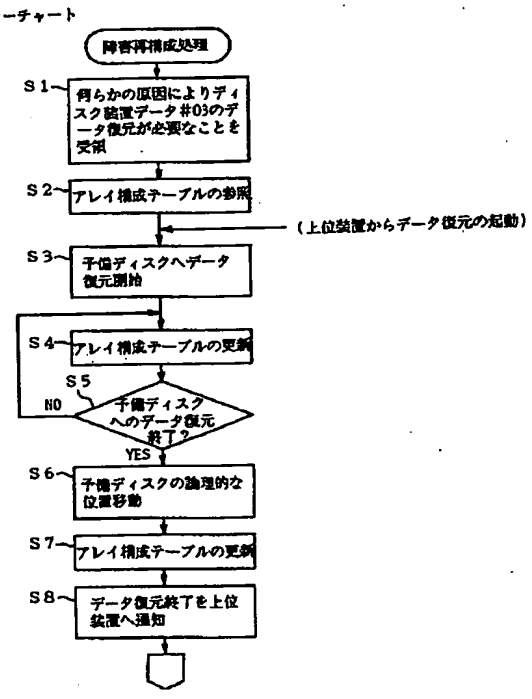
	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02		パリティ#0	データ#03
	00D	01D	02D	03P	04P	05D

(F)（ディスク移動後）

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	
	00D	01D	02D	03D	04P	

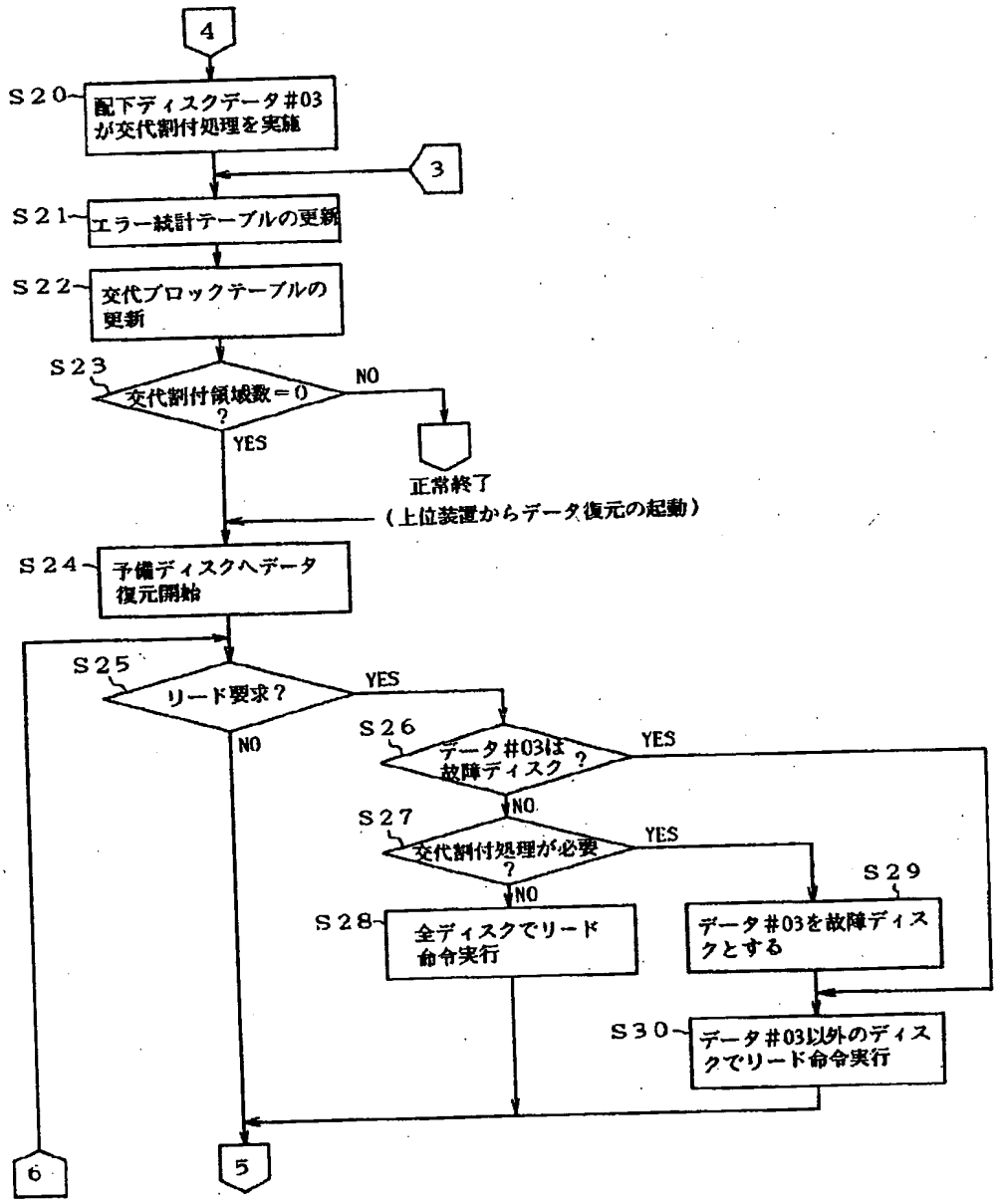
【図１９】

人的介入により物理的なディスク装置の移動を必要としない再構成処理を示したフローチャート



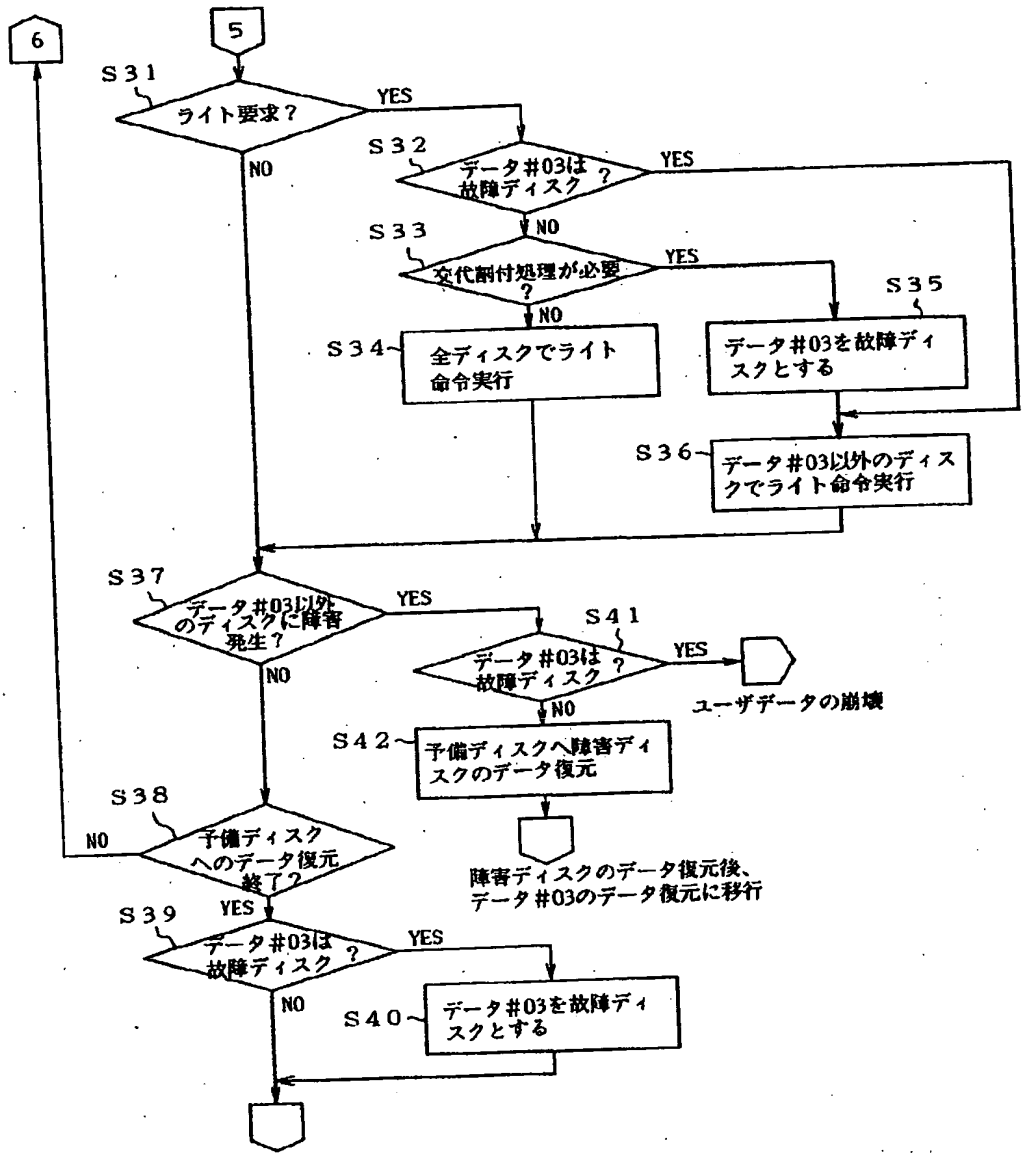
【図5】

交代領域オーバフローに伴う予備へのデータ復元処理を示したフローチャート



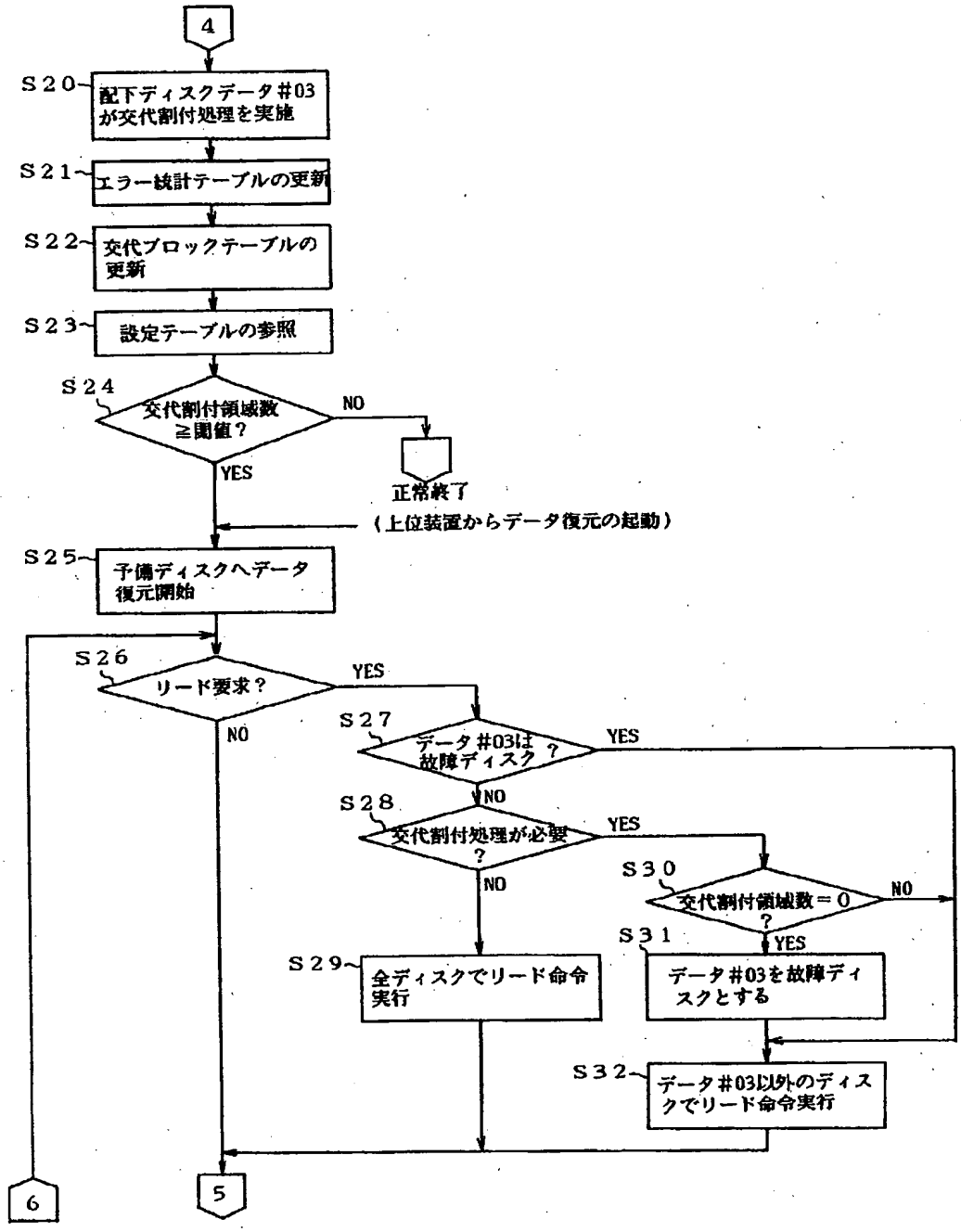
【図６】

図５の続きを示したフローチャート



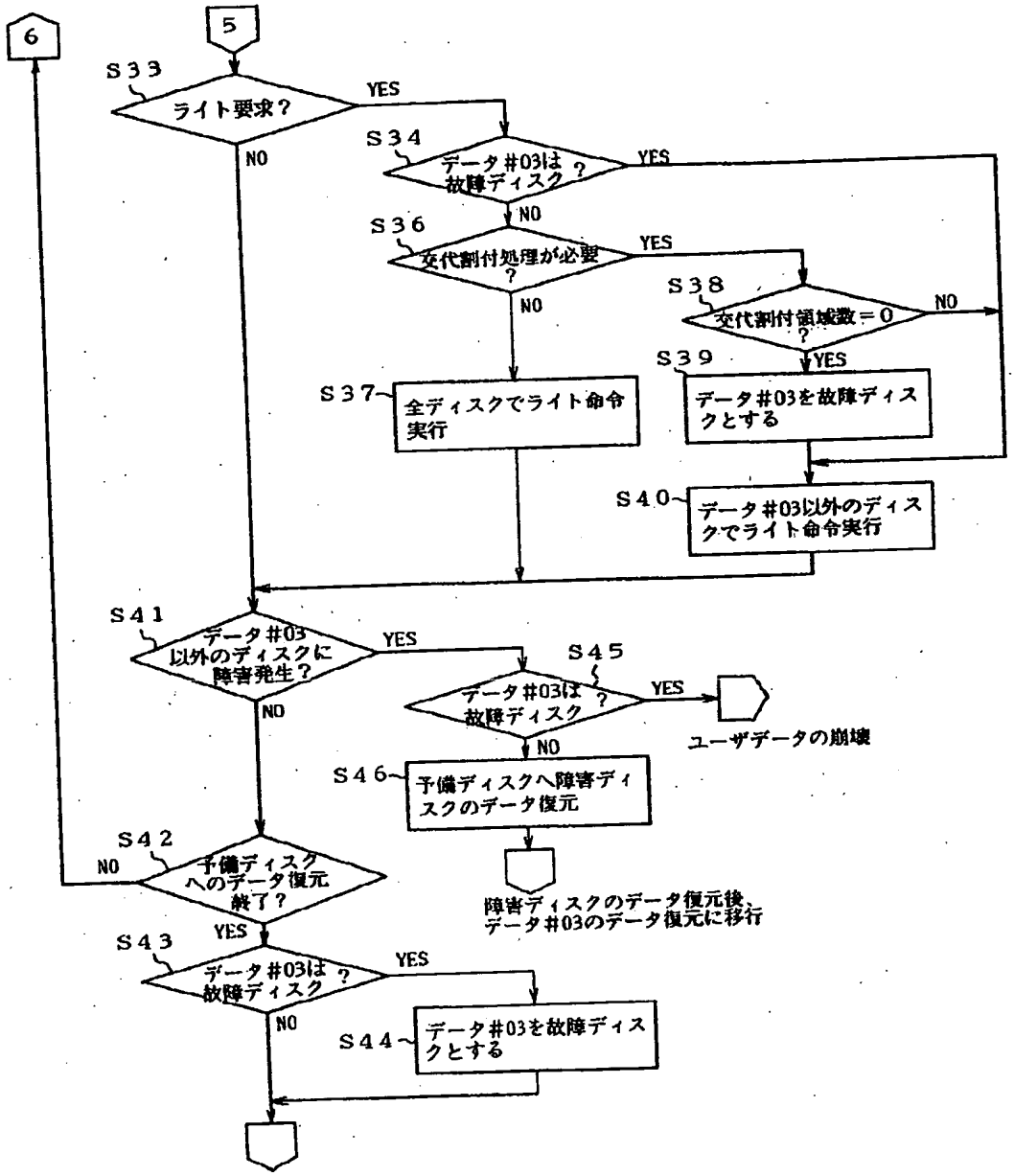
【図７】

交代領域オーバフローの予測に基づくデータ復元処理を示したフローチャート



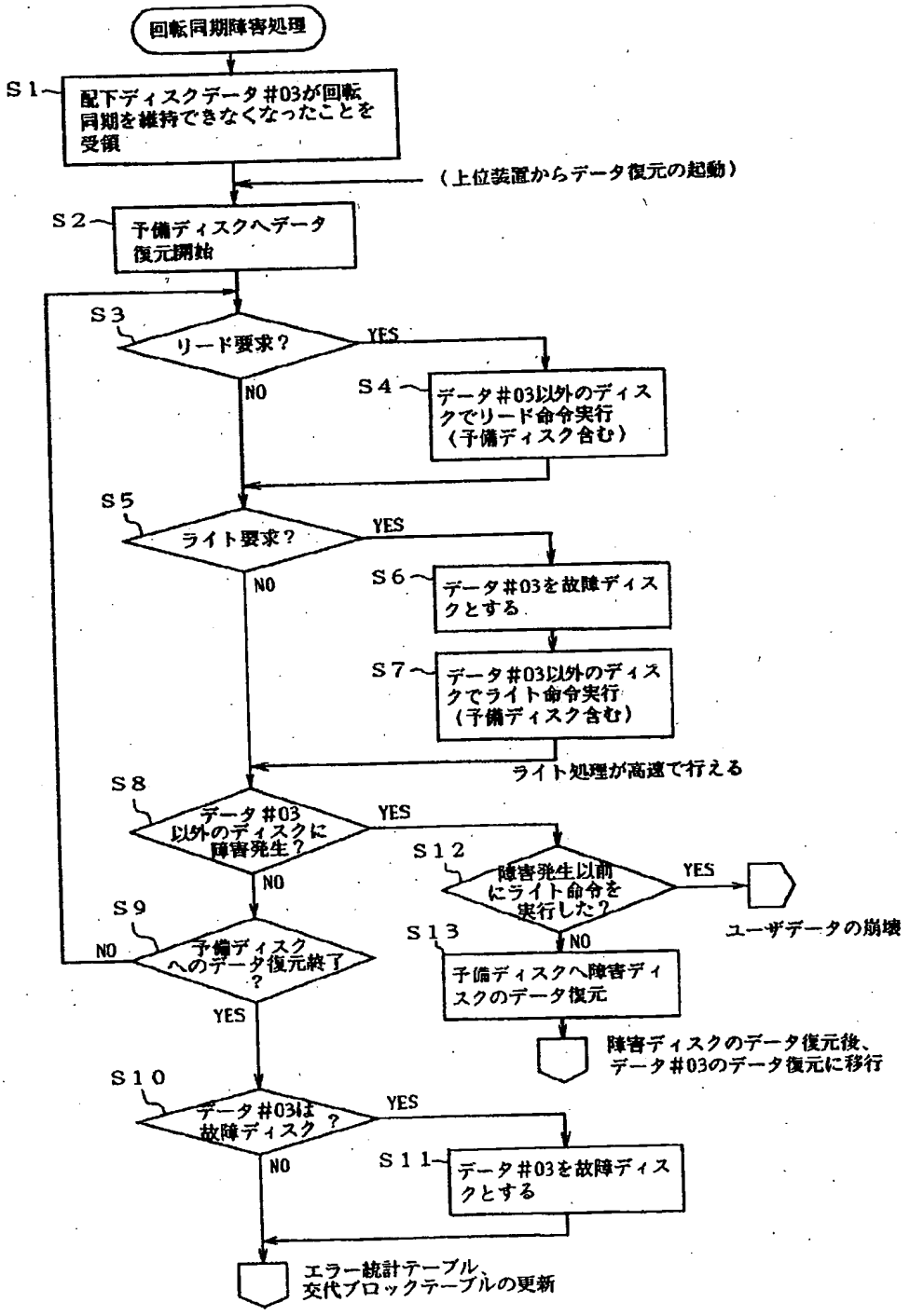
【図８】

図７の続きを示したフローチャート



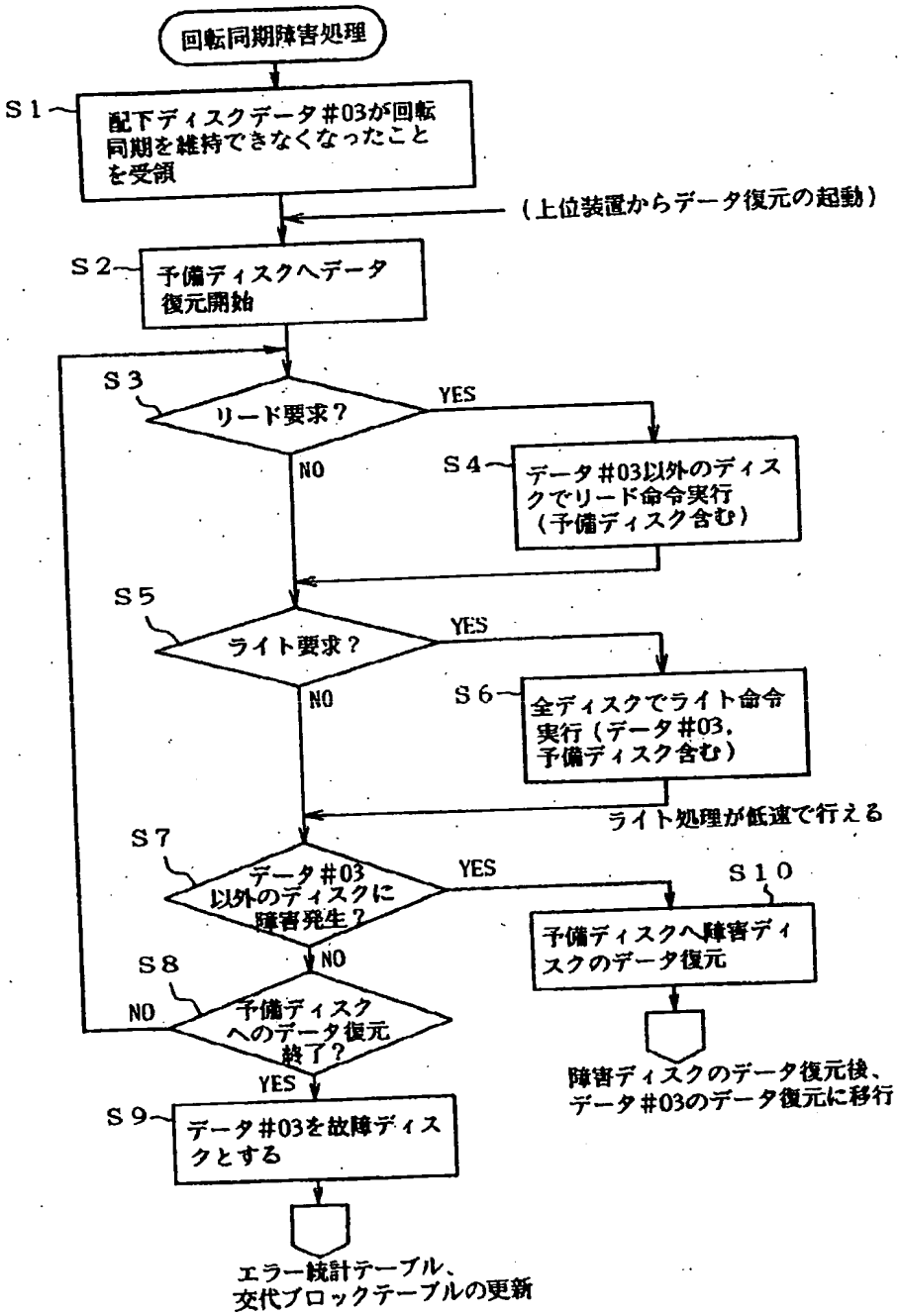
【図９】

同期回転異常時の予備へのデータ復元処理を示したフローチャート



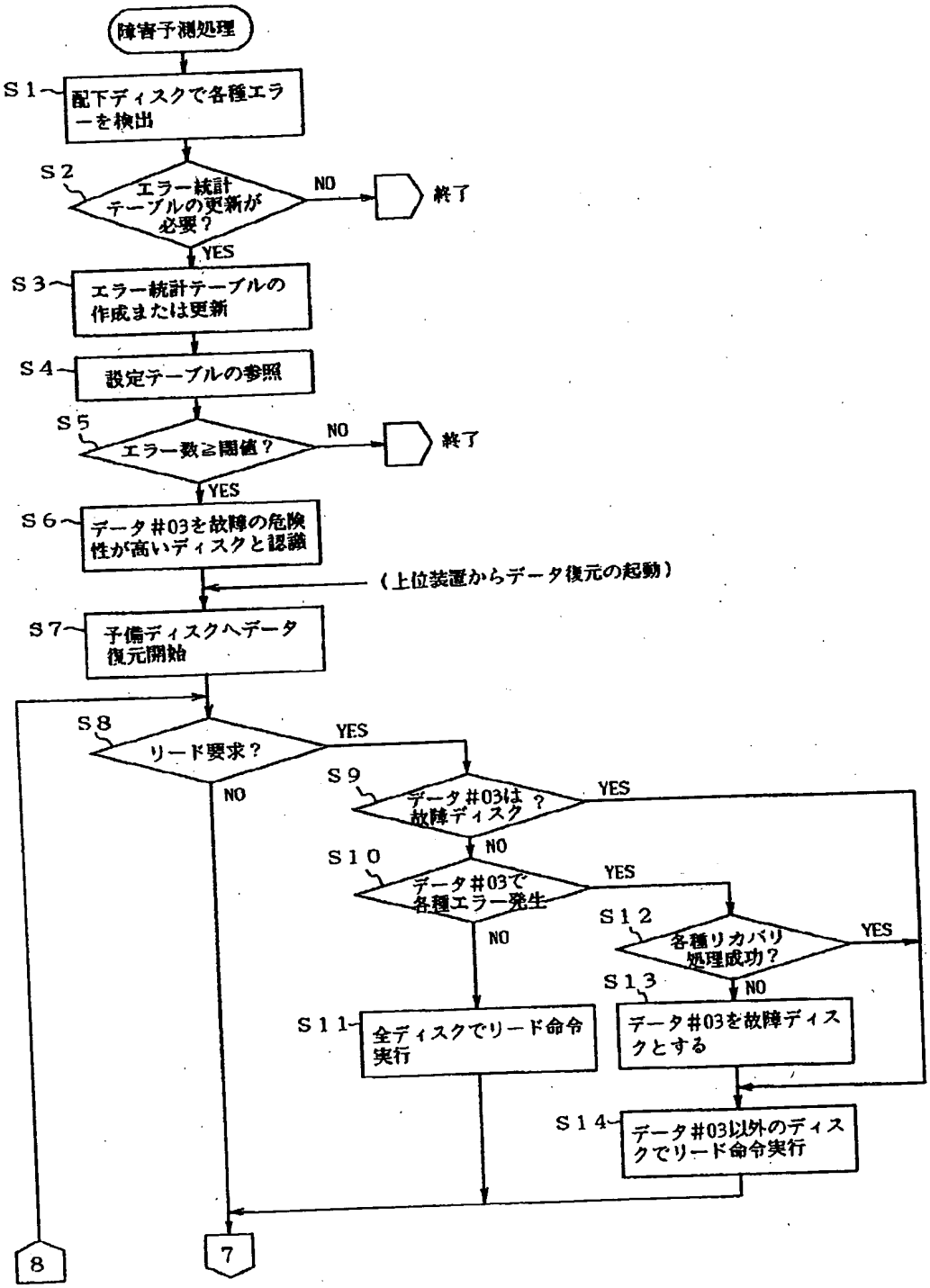
【図１０】

同期回転異常時の予備へのデータ復元処理の他の実施例を示したフローチャート



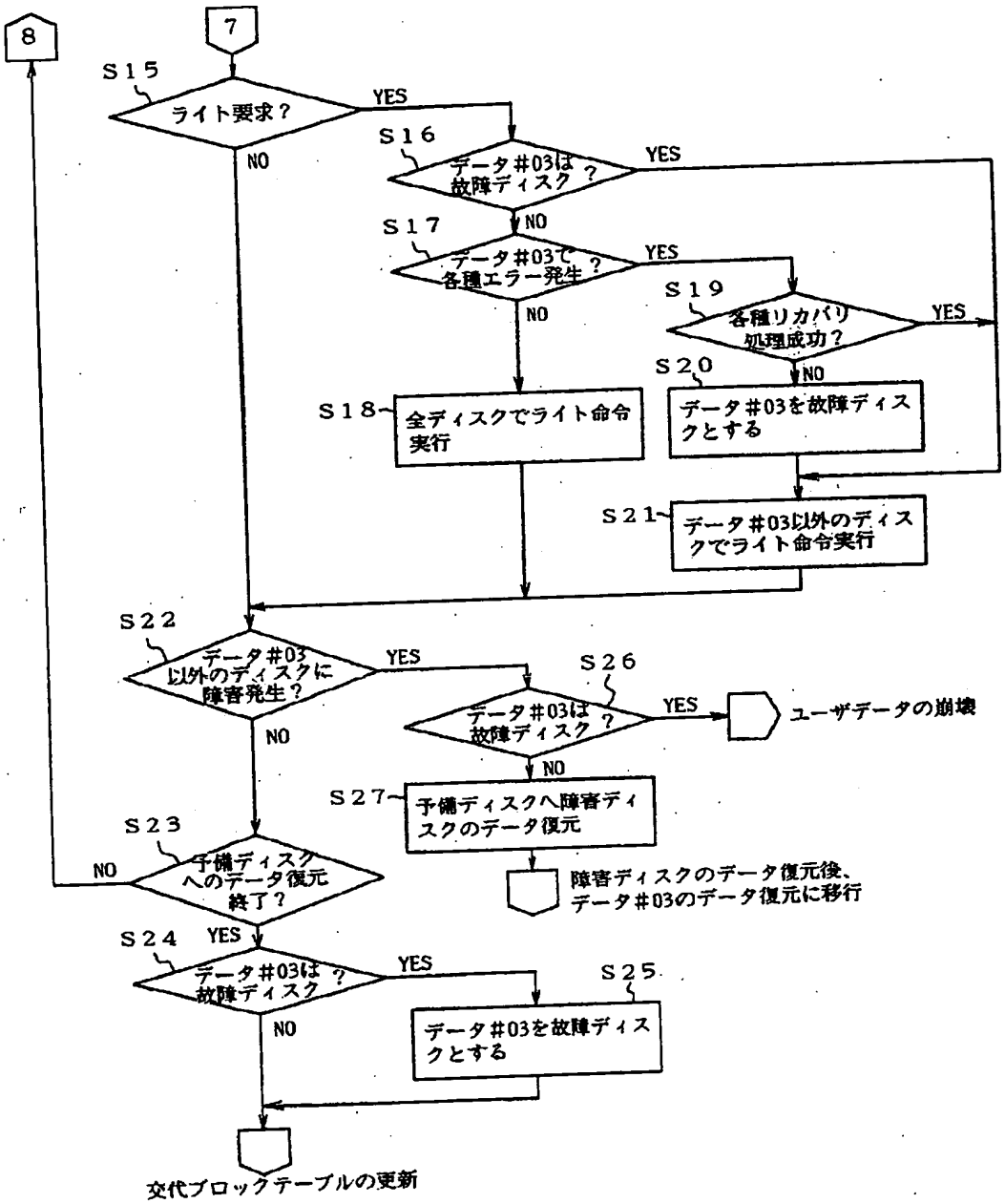
【図１１】

障害発生の予測と予備へのデータ復元処理を示したフローチャート



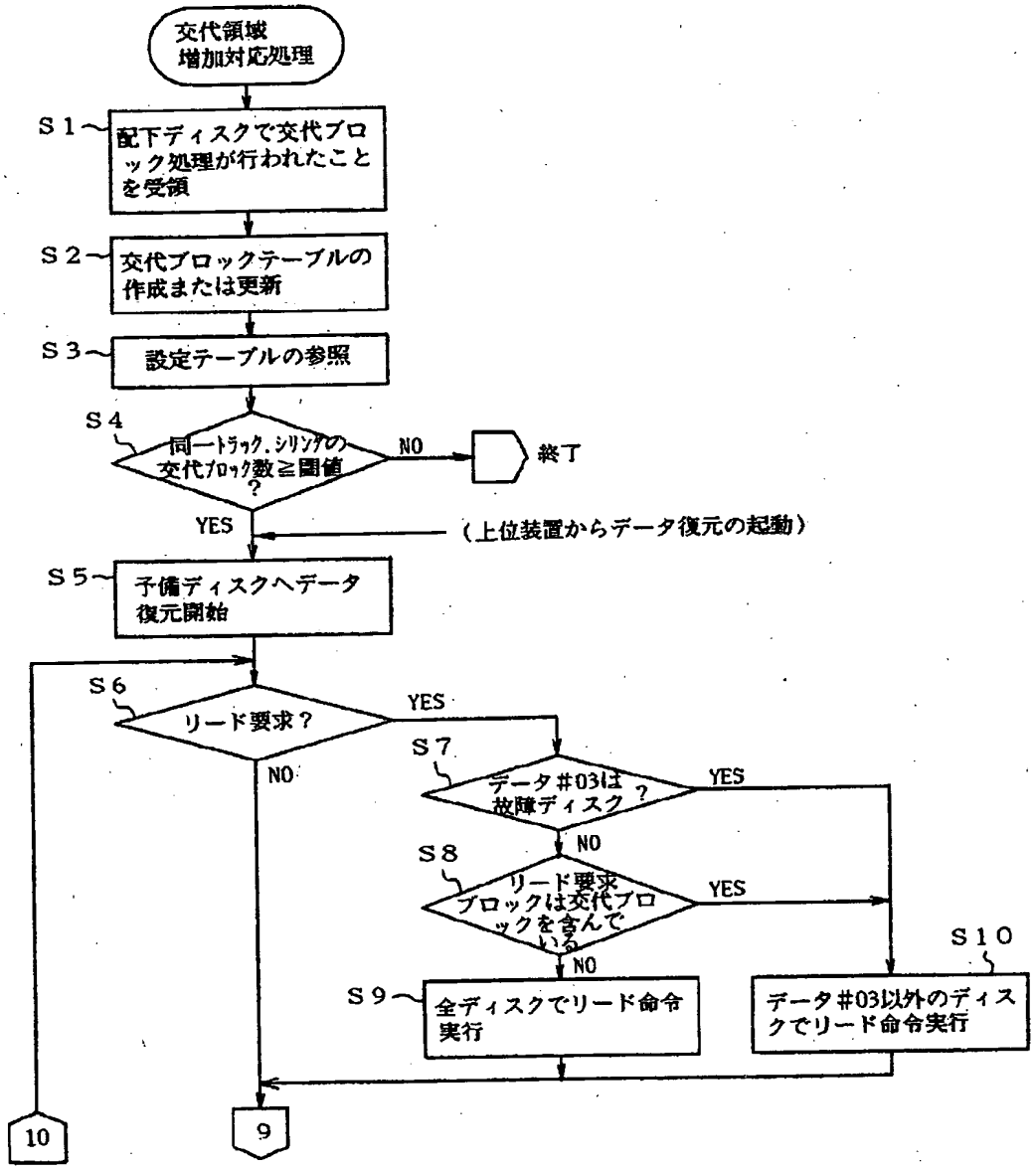
【図１２】

図１１の続きを示したフローチャート



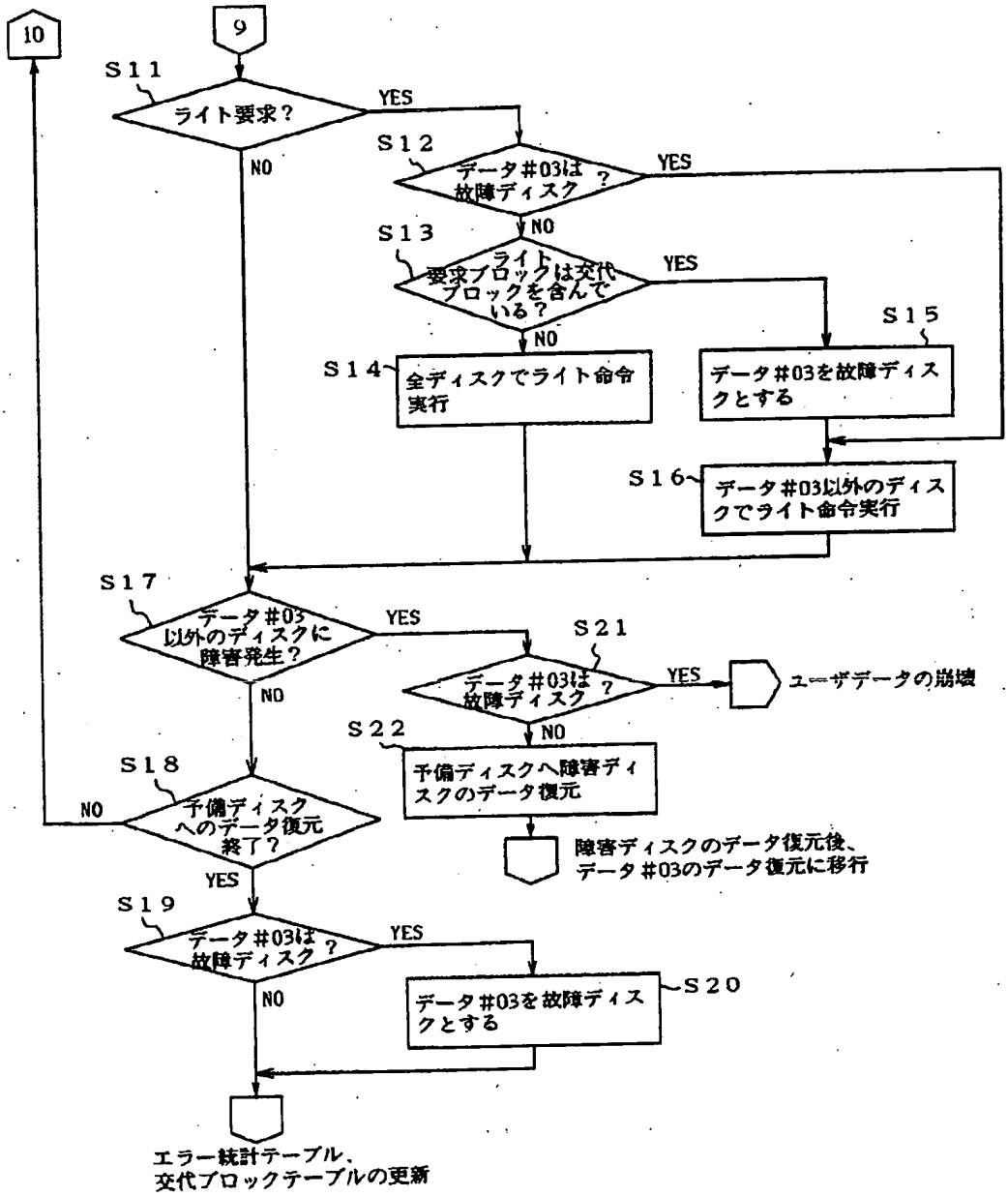
【図１３】

同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理を示したフローチャート



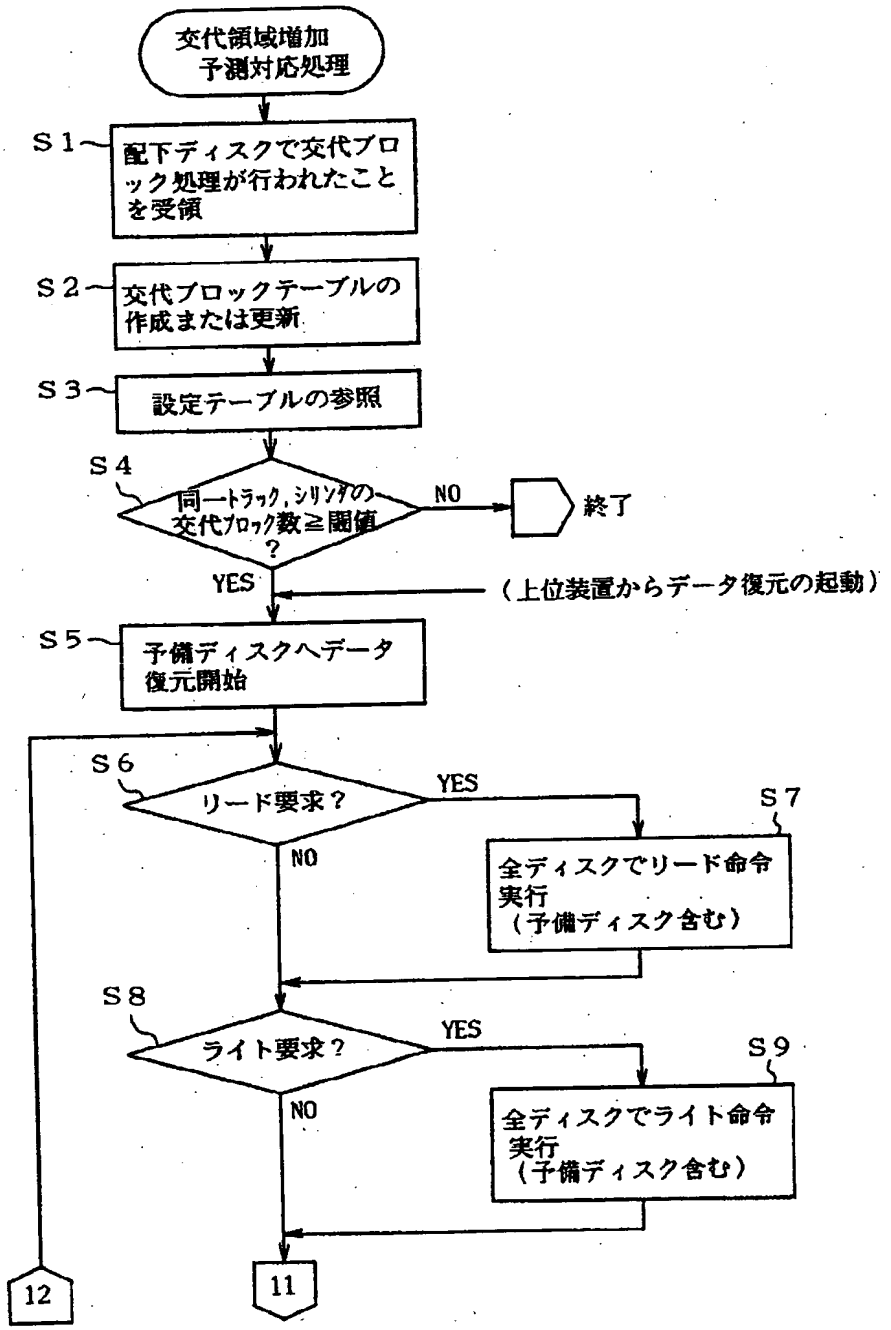
【図１４】

図１３の続きを示したフローチャート



【図１５】

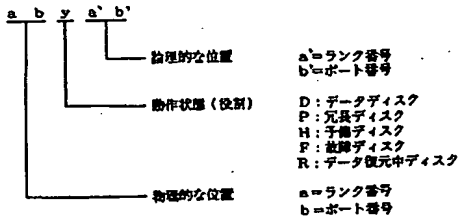
同一トラック又はシリンダ内の交代ブロック数が増加した場合の予備へのデータ復元処理の他の実施例を示したフローチャート



【図２０】

図１９の処理で更新されるアレイ構成テーブルの内容を示した説明図

(A)



(データ復元前)

(B)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	ホットスペア#0
	00D00	01D01	02D02	03D03	04P04	05H05

(データ復元中)

(C)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02	データ#03	パリティ#0	ホットスペア#0
	00D00	01D01	02D02	03F03	04P04	05R05

(データ復元後)

(D)

	ポート#0	ポート#1	ポート#2	ポート#3	ポート#4	ポート#5
ランク#0	データ#00	データ#01	データ#02		パリティ#0	データ#03
	00D00	01D01	02D02	03F03	04P04	05D03

【図２１】

従来装置の説明図